

CERTIFICATE OF MAILING

I hereby certify that this paper or, if this paper is a transmittal letter, every other paper or fee referred to therein, is being deposited with the U.S. Postal Service as first class mail in an envelope addressed to Commissioner of Patents & Trademarks, Washington, DC 20231, on

(Date of Deposit)

Date

Name

PLEASE CHARGE ANY DEFICIENCY UP TO \$300.00 OR CREDIT ANY EXCESS IN THE FEES DUE WITH THIS DOCUMENT TO OUR DEPOSIT ACCOUNT NO. 04-0100

Docket No. 1034/1F808-US7

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Peter H. St. George-Hyslop; Johanna M. Rommens; and Paul E. Fraser

For: GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE

CONTINUING APPLICATION

Hon. Commissioner of
Patents and Trademarks
Washington, DC 20231

Sir:

This is a request for filing under 37 C.F.R. 1.53(b) of a:

☐ Continuation ☒ Divisional

application of the following pending prior application:

Serial No. 08/509,359

Filed: July 31, 1995

Of: Peter H. St. George-Hyslop; Johanna M. Rommens; and Paul E. Fraser

For: GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE

Examiner: J. Burke

Group: 1642

1. Enclosed is a copy of the prior application as originally filed (along with a copy of the original Declaration). No amendments identified in the

EXPRESS MAIL CERTIFICATE

10/12/00 62822300 4US

I hereby certify that, on the date indicated above I deposited this paper or fee with the U.S. Postal Service & that it was addressed for delivery to the Commissioner of Patents & Trademarks, Washington D.C. 20231 by "Express Mail Post Office to Addressee" service.

Name (Print)

Signature

JC925 U.S. PTO
09/689159

10/12/00

declaration for this prior application introduced new matter. Please use this copy as the application and declaration for the present case.

2. The filing fee is calculated below:

CLAIMS AS FILED, AFTER ACCOMPANYING AMENDMENT

	Claims on File	Number Extra	Rate
Basic Fee			\$710.00
Total Claims	72 - 20 =	52 x \$18	\$0.00
Independent Claims	28 - 3 =	25 x \$78	\$0.00
If Multiple Dependent Claims Are Present, Add \$260.00			\$0.00
Total Filing Fee			\$0.00
For Small Entity (half of preceding total)*			\$

*No. 9 below must be checked to claim this reduction.

3. A check in the amount of \$0.00 is enclosed.
4. ☐ Cancel claims .
5. ☒ Amend the specification by inserting before the first line the sentence (check one and fill in):
- "This is a ☐ continuation, ☒ division, of application Serial No. 08/509,359, filed July 31, 1995. Each of these prior applications is hereby incorporated herein by reference, in its entirety."
6. ☒ The prior application is assigned to: The Governing Council of the University of Toronto and

HSC Research and Development
Limited Partnership

7. ☐ A Preliminary Amendment is also enclosed.
8. ☒ Informal Drawings are filed herewith
9. ☐ A verified statement claiming small entity status (check one):
☐ was filed in parent application or ☐ is enclosed.
10. ☐ Priority is claimed from
Country:
Number:
Date:

The priority document

- ☐ was filed in the prior application
☐ is enclosed.

11. The Assignment
☒ was recorded at Reel 8008, Frame 0101 on June 27, 1996
☐ is enclosed.
12. Copy of Associate Power of Attorney (and change of correspondence address) from parent application.

Please direct all correspondence to the undersigned.

Respectfully submitted,



Paul F. Fehlner, Ph.D.

Reg. No. 35,135

Attorney of Record in Prior Application

DARBY & DARBY P.C.
805 Third Avenue
New York, NY 10022
212-527-7700

10/12/00 62822300405

I hereby certify that, on the date indicated above I deposited this paper or fee with the U.S. Postal Service & that it was addressed for delivery to the Commissioner of Patents & Trademarks, Washington, D.C. 20531 by Express Mail Post Office to Addressee Service.

PRESS MAIL NO. TB879465048

TITLE OF THE INVENTION

Name (Print)

Signature

**GENETIC SEQUENCES AND PROTEINS RELATED TO
ALZHEIMER'S DISEASE**5 RELATED APPLICATIONS

This application is a Continuation-In-Part of U.S. application Serial No. _____, entitled GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE (Inventors: Peter H. St. George-Hyslop, Johanna N. Rommens and Paul E. Fraser), filed on June 28, 1995, which was a Continuation-in-Part of U.S. Application Serial No. 08/431,048, filed April 28, 1995.

10 FIELD OF THE INVENTION

The present invention relates generally to the field of neurological and physiological dysfunctions associated with Alzheimer's Disease. More particularly, the invention is concerned with the identification, isolation and cloning of the gene which when mutated is associated with Alzheimer's Disease as well as its transcript, gene products and associated sequence information and neighbouring genes. The present invention also relates to methods of diagnosing for and detection of carriers of the gene, Alzheimer's Disease diagnosis, gene therapy using recombinant technologies and therapy using the information derived from the DNA, protein, and the metabolic function of the protein.

20 BACKGROUND OF THE INVENTION

In order to facilitate reference to various journal articles, a listing of the articles is provided at the end of this specification.

Alzheimer's Disease (AD) is a degenerative disorder of the human central nervous system characterized by progressive memory impairment and cognitive and intellectual decline during mid to late adult life (Katzman, 1986). The disease is accompanied by a constellation of neuropathologic features principal amongst which are the presence of extracellular amyloid or senile plaques and the neurofibrillary degeneration of neurons. The etiology of this disease is complex, although in some

families it appears to be inherited as an autosomal dominant trait. However, even amongst these inherited forms of AD, there are at least three different genes which confer inherited susceptibility to this disease (St George-Hyslop et al., 1990). The $\epsilon 4$ (Cys112Arg) allelic polymorphism of the Apolipoprotein E (*ApoE*) gene has been associated with AD in a significant proportion of cases with onset late in life (Saunders et al., 1993; Strittmatter et al., 1993). Similarly, a very small proportion of familial cases with onset before age 65 years have been associated with mutations in the β -amyloid precursor protein (*APP*) gene (Chartier-Harlin et al., 1991; Goate et al., 1991; Murrell et al., 1991; Karlinsky et al., 1992; Mullan et al., 1992). A third locus (AD3) associated with a larger proportion of cases with early onset AD has recently been mapped to chromosome 14q24.3 (Schellenberg et al., 1992; St George-Hyslop et al., 1992; Van Broeckhoven et al., 1992).

Although chromosome 14q carries several genes which could be regarded as candidate genes for the site of mutations associated with AD3 (e.g. *cFOS*, α -1-antichymotrypsin, and cathepsin G), most of these candidate genes have been excluded on the basis of their physical location outside the AD3 region and/or the absence of mutations in their respective open reading frames (Schellenberg, GD et al., 1992; Van Broeckhoven, C et al., 1992; Rogaev et al., 1993; Wong et al., 1993).

There have been several developments and commercial directions in respect of treatment of Alzheimer's Disease and diagnosis thereof. Published PCT application WO 94 23049 describes transfection of high molecular weight YAC DNA into specific mouse cells. This method is used to analyze large gene complexes, for example the transgenic mice may have increased amyloid precursor protein gene dosage, which mimics the trisomic condition that prevails in Down's Syndrome and the generation of animal models with β -amyloidosis prevalent in individuals with Alzheimer's Disease. Published international application WO 94 00569 describes transgenic non-human animals harbouring large trans genes such as the trans gene comprising a human amyloid precursor protein gene. Such animal models can provide useful models of human genetic diseases such as Alzheimer's Disease.

Canadian Patent application 2096911 describes a nucleic acid coding for amyloid precursor protein-cleaving protease, which is associated with Alzheimer's Disease and Down's syndrome. The genetic information may be used to diagnose Alzheimer's disease. The genetic information was isolated from chromosome 19.

5 Canadian patent application 2071105, describes detection and treatment of inherited or acquired Alzheimer's disease by the use of YAC nucleotide sequences. The YACs are identified by the numbers 23CB10, 28CA12 and 26FF3.

U.S. Patent 5297562, describes detection of Alzheimer's Disease having two or more copies of chromosome 21. Treatment involves methods for reducing the proliferation of chromosome 21 trisomy. Canadian Patent application 2054302, describes monoclonal antibodies which recognize human brain cell nucleus protein encoded by chromosome 21 and are used to detect changes or expression due to Alzheimer's Disease or Down's Syndrome. The monoclonal antibody is specific to a protein encoded by human chromosome 21 and is linked to large pyramidal cells of human brain tissue.

10

15

By extensive effort and a unique approach to investigating the AD3 region of chromosome 14q, the Alzheimer's related membrane protein (ARMP) gene has been isolated, cloned and sequenced from within the AD3 region on chromosome 14q24.3. In addition, direct sequencing of RT-PCR products spanning this 3.0 kb cDNA transcript isolated from affected members of at least eight large pedigrees linked to chromosome 14, has led to the discovery of missense mutations in each of these different pedigrees. These mutations are absent in normal chromosomes. It has now been established that the ARMP gene is causative of familial Alzheimer's Disease type AD3. In realizing this link, it is understood that mutations in this gene can be associated with other cognitive, intellectual, or psychological diseases such as cerebral hemorrhage, schizophrenia, depression, mental retardation and epilepsy. These phenotypes are present in these AD families and these phenotypes have been seen in mutations of the APP protein gene. The Amyloid Precursor Protein (APP) gene is also associated with inherited Alzheimer's Disease. The identification of both normal and mutant forms of the ARMP gene and gene products has allowed for the

20

25

30

development of screening and diagnostic tests for ARMP utilizing nucleic acid probes and antibodies to the gene product. Through interaction with the defective gene product and the pathway in which this gene product is involved, gene therapy, manipulation and delivery are now made possible.

5

SUMMARY OF THE INVENTION

Various aspects of the invention are summarized as follows. In accordance with a first aspect of the invention, a purified mammalian polynucleotide is provided which codes for Alzheimer's related membrane protein (ARMP). The polynucleotide has a sequence which is the functional equivalent of the DNA sequence of ATCC deposit _____, deposited April 28, 1995. The mammalian polynucleotide may be in the form of DNA, genomic DNA, cDNA, mRNA and various fragments and portions of the gene sequence encoding ARMP. The mammalian DNA is conserved in many species, including humans and rodents, example mice. The mouse sequence encoding ARMP has greater than 95% homology with the human sequence encoding the same protein.

Purified human nucleotide sequences which encode mutant ARMP have mutations at nucleotide position i) 685, A→C ii) 737, A→G iii) 986, C→A, iv) 1105, C→G, v) 1478, G→A, vi) 1027, C→T, vii) 1102, C→T and viii) 1422, C→G of Sequence ID No: 1 as well as in the cDNA sequence of a further human clone of a sequence identified by ID NO:132.

The nucleotide sequences encoding ARMP have an alternative splice form in the genes open reading frame. The human cDNA sequence which codes for ARMP has sequence ID No. 1 as well as sequence ID NO:132¹³³ as sequenced in a another human clone. The mouse sequence which encodes ARMP has sequence ID No. 3, as well as SEQ ID NO:134¹³⁵ derived from a further clone containing the entire coding region. Various DNA and RNA probes and primers may be made from appropriate polynucleotide lengths selected from the sequences. Portions of the sequence also encode antigenic determinants of the ARMP.

Suitable expression vectors comprising the nucleotide sequences are provided

along with suitable host cells transfected with such expression vectors.

In accordance with another aspect of the invention, purified mammalian Alzheimer's related membrane protein is provided. The purified protein has an amino acid sequence encoded by polynucleotide sequence as identified above which for the human is sequence ID NO:2 and SEQ ID NO:133 (derived from another clone).
The mouse amino acid sequence is defined by sequence ID No. 2 and sequence ID No. 4, the later being translated from another clone containing the entire coding region. The purified protein may have substitution mutations selected from the group consisting of positions identified in Sequence ID No: 2 and Sequence ID NO:133.

- 10 i) M 146L
- ii) H 163R
- iii) A 246E
- iv) L 286V
- v) C 410 Y
- 15 vi) A 260 V
- vii) A 285 V
- viii) L 392 V

In accordance with another aspect of the invention, are polyclonal antibodies raised to specific predicted sequences of the ARMP protein. Polypeptides of at least six amino acid residues are provided. The polypeptides of six or greater amino acid residues may define antigenic epitopes of the ARMP. Monoclonal antibodies having suitably specific binding affinity for the antigenic regions of the ARMP are prepared by use of corresponding hybridoma cell lines. In addition, other polyclonal antibodies may be prepared by inoculation of animals with suitable peptides or holoprotein which add suitable specific binding affinities for antigenic regions of the ARMP.

In accordance with another aspect of the invention, an isolated DNA molecule is provided which codes for E5-1 protein.

In accordance with another aspect of the invention, purified E5-1 protein is provided, having amino acid Sequence ID No: 137.

In accordance with another aspect of the invention a bioassay is provided for

determining if a subject has a normal or mutant ARMP, where the bioassay comprises providing a biological sample from the subject

conducting a biological assay on the sample to detect a normal or mutant gene sequence coding for ARMP, a normal or mutant ARMP amino acid sequence, or a normal or defective protein function.

In accordance with another aspect of the invention, a process is provided for producing ARMP comprising culturing one of the above described transfected host cells under suitable conditions, to produce the ARMP by expressing the DNA sequence. Alternatively, ARMP may be isolated from mammalian cells in which the ARMP is normally expressed.

In accordance with another aspect of the invention, is a therapeutic composition comprising ARMP and a pharmaceutically acceptable carrier.

In accordance with another aspect of the invention, a recombinant vector for transforming a mammalian tissue cell to express therapeutically effective amounts of ARMP in the cells is provided. The vector is normally delivered to the cells by a suitable vehicle. Suitable vehicles include vaccinia virus, adenovirus, adeno associated virus, retrovirus, liposome transport, neurotropic viruses, Herpes simplex virus and other vector systems.

In accordance with another aspect of the invention, a method of treating a patient deficient in normal ARMP comprising administering to the patient a therapeutically effective amount of the protein targeted at a variety of patient cells which normally express ARMP. The extent of administration of normal ARMP being sufficient to override any effect the presence of the mutant ARMP may have on the patient. As an alternative to protein, suitable ligands and therapeutic agents such as small molecules and other drug agents may be suitable for drug therapy designed to replace the protein and defective ARMP, displace mutant ARMP, or to suppress its formation.

In accordance with another aspect of the invention an immuno therapy for treating a patient having Alzheimer's Disease comprises treating the patient with antibodies specific to the mutant ARMP to reduce biological levels or activity of the

mutant ARMP in the patient. To facilitate such amino acid therapy, a vaccine composition may be provided for evoking an immune response in a patient of Alzheimer's Disease where the composition comprises a mutant ARMP and a pharmaceutically acceptable carrier with or without a suitable excipient. The antibodies developed specific to the mutant ARMP could be used to target appropriately encapsulated drugs/molecules, specific cellular/tissue sites. Therapies utilizing specific ligands which bind to normal or wild type ARMP of either mutant or wild type and which augments normal function of ARMP in membranes and/or cells or inhibits the deleterious effect of the mutant protein are also made possible.

In accordance with another aspect of the invention, a transgenic animal model for Alzheimer's Disease which has the mammalian polynucleotide sequence with at least one mutation which when expressed results in mutant ARMP in the animal cells and thereby manifests a phenotype. For example, the human Prion gene when over-expressed in rodent peripheral nervous system and muscle cells causes a quite different response in the animal than the human. The animal may be a rodent and is preferably a mouse, but may also be other animals including rat, pig, *Irosophila melanogaster*, *C. elegans* (nematode), all of which are used for transgenic models. Yeast cells can also be used in which the ARMP Sequence is expressed from an artificial vector.

In accordance with another aspect of the invention a transgenic mouse model for Alzheimer's Disease has the mouse gene encoding ARMP human or murine homologues mutated to manifest the symptoms. The transgenic mouse may exhibit symptoms of cognitive memory or behavioural disturbances. In addition or alternatively, the symptoms may appear as another cellular tissue disorders such as in mouse liver, kidney spleen or bone marrow and other organs in which the ARMP gene product is normally expressed.

In accordance with another aspect of the invention, the protein can be used as a starting point for rationale drug design to provide ligands, therapeutic drugs or other types of small chemical molecules.

BRIEF DESCRIPTION OF THE DRAWINGS

Various aspects of the invention are described hereinafter with respect to the drawings wherein:

5 Figure 1a. Genomic physical and transcriptional map of the AD3 region of chromosome 14. Genetic map inter-marker genetic distances averaged for male and female meiosis are indicated in centiMorgans.

10 Figure 1b. Is the constructed physical contig map of overlapping genomic DNA fragments cloned into YACs spanning a FAD locus on chromosome 14q.

Figure 1c. Regions of interest within the constructed physical contig map.

15 Figure 1d. Transcriptional map illustrating physical locations of the 19 independent longer cDNA clones.

Figure 2. Automated fluorescent chromatograms representing the change in nucleic acids which direct (by the codon) the amino acid sequence of the gene.

- 20 (a) Met 146 Leu
 (b) His 163 Arg
 (c) Ala 246 Glu
 (d) Leu 286 Val
 (e) Cys 410 Tyr

25 Figure 3(a). Restriction fragments of M 146 L mutation using BspHI restriction enzyme in AD patients. Absence of a restriction site indicates a mutant allele.

Figure 3(b). Presence of the His 163 Arg mutation detected by NlaIII restriction digestion. Absence of a restriction indicates a mutant allele.

30

Figure 3(c). Presence of the Ala 246 Glu mutation in AD patients using DdeI restriction enzyme. Presence of mutant allele leads to restriction.

5 Figure 3(d). Presence of Cys 410 Tyr mutation in AD patients as assayed using allele specific oligonucleotides.

Figure 3(e). Presence of Leu286Val mutation in AD patients using PvuII restriction enzyme in AD patients.

10 Figure 4. RNA blot demonstrating the expression of ARMP protein mRNA in different regions of the brain including amygdala, caudate, corpus callosum, hippocampus, hypothalamus, substantia nigra, subthalamic nucleus and thalamus.

15 Figure 5. RNA blot demonstrating the expression of ARMP protein mRNA in a variety of tissues including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas.

Figure 6a. Hydropathy plot of the putative ARMP protein.

20 Figure 6b. A model for the structural organization of the putative ARMP protein. Roman numerals depict the transmembrane domains. Putative glycosylation sites are indicated as asterisks and most of the phosphorylation sites are located on the same membrane face as the two acidic hydrophilic loops. The MAP kinase site is present at residue 115 and the PKC site at residue 114. FAD mutation sites are indicated by
25 horizontal arrows.

Figure 7 shows transcription of the *E5-1* gene, investigated by hybridization of the *E5-1* cDNA to Northern blots of mRNA from multiple human brain regions (Panel A), and several peripheral tissues (Panel C). In brain, the *E5-1* transcript is of a
30 lower molecular weight and lesser abundance than the ARMP transcript (Panel B)

hybridized to the same blot using identical conditions.

Figure 8 shows the predicted structure of the E5-1 protein.

5 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

In order to facilitate review of the various embodiments of the invention and an understanding of various elements and constituents used in making the invention and using same, the following definition of terms used in the invention description is as follows:

10 **Alzheimer Related Membrane Protein gene (ARMP gene)** - the chromosome 14 gene which when mutated is associated with familial Alzheimer's Disease and/or other inheritable disease phenotypes (eg. cerebral hemorrhage, mental retardation, schizophrenia, psychosis, and depression). This definition is understood to include the various sequence polymorphisms that exist, wherein nucleotide substitutions in the
15 gene sequence do not affect the essential function of the gene product, as well as functional equivalents of the nucleotide sequences of Sequence ID No. 1, Sequence ID NO:132, Sequence ID No: 3 and Sequence ID NO:134. This term primarily relates to an isolated coding sequence, but can also include some or all of the flanking regulatory elements and/or introns. The term ARMP gene includes the gene in other
20 species analogous to the human gene which when mutated is associated with Alzheimer's disease.

Alzheimer Related Membrane Protein (ARMP) - the protein encoded by the ARMP gene. The preferred source of protein is the mammalian protein as isolated from humans or animals. Alternatively, functionally equivalent proteins may exist in
25 plants, insects and invertebrates (such as *C. elegans*). The protein may be produced by recombinant organisms, or chemically or enzymatically synthesized. This definition is understood to include functional variants such as the various polymorphic forms of the protein wherein amino acid substitutions or deletions within the amino acid sequence do not affect the essential functioning of the protein, or its structure.
30 It also includes functional fragments of ARMP.

Mutant ARMP gene - The ARMP gene containing one or more mutations which lead to Alzheimer's Disease and/or other inheritable disease phenotypes (eg. cerebral hemorrhage, mental retardation, schizophrenia, psychosis, and depression). This definition is understood to include the various mutations that exist, wherein nucleotide substitutions in the gene sequence affect the essential function of the gene product, as well as mutations of functional equivalents of the nucleotide sequences of Sequence ID No. 1, Sequence ID NO:132, Sequence ID No:3, and ID NO:134 (the corresponding amino acid sequences). This term primarily relates to an isolated coding sequence, but can also include some or all of the flanking regulatory elements and/or introns.

Mutant ARMP - a mammalian protein that is highly analogous to ARMP in terms of primary structure, but wherein one or more amino acid deletions and/or substitutions result in impairment of its essential function, so that mammals, especially humans, whose ARMP producing cells express mutant ARMP rather than the normal ARMP, demonstrate the symptoms of Alzheimer's Disease and/or other relevant inheritable phenotypes (eg. cerebral hemorrhage, mental retardation, schizophrenia, psychosis, and depression).

mARMP gene - mouse gene analogous to the human ARMP gene. Functional equivalent as used in describing gene sequences and amino acid sequences means that a recited sequence need not be identical to the definitive sequence of the Sequence ID Nos but need only provide a sequence which functions biologically and/or chemically the equivalent of the definitive sequence. Hence sequences which correspond to a definitive sequence may also be considered as functionally equivalent sequence.

mARMP - mouse Alzheimer related membrane protein, analogous to the human ARMP, encoded by the mARMP gene. This definition is understood to include the various polymorphic forms of the protein wherein amino acid substitutions or deletions of the sequence does not affect the essential functioning of the protein, or its structure.

Mutant mARMP - a mouse protein which is highly analogous to mARMP in terms of primary structure, but wherein one or more amino acid deletions and/or

substitutions result in impairment of its essential function, so that mice, whose mARMP producing cells express mutant mARMP rather than the normal mARMP demonstrate the symptoms of Alzheimer's Disease and/or other relevant inheritable phenotypes, or other phenotypes and behaviours as manifested in mice.

5 **ARMP carrier** - a mammal in apparent good health whose chromosomes contain a mutant ARMP gene that may be transmitted to the offspring and who will develop Alzheimer's Disease in mid to late adult life.

Missense mutation - A mutation of nucleic acid sequence which alters a codon to that of another amino acid, causing an altered translation product to be made.

10 **Pedigree** - In human genetics, a diagram showing the ancestral relationships and transmission of genetic traits over several generations in a family.

E5-1 gene - the chromosome 1 gene which shows homology to the ARMP gene and which when mutated is associated with familial Alzheimer's disease and/or other inheritable disease phenotypes. This definition is understood to include the various sequence polymorphisms that exist, wherein nucleotide substitutions in the gene sequence do not affect the essential function of the gene product, as well as functional equivalents of the nucleotide Sequence ID No: 136. This term also includes the gene in other species analogous to the human gene described herein.

20 **E5-1 protein** - the protein encoded by the *E5-1* gene. This term includes the protein of Sequence ID No: 137 and also functional variants such as the various polymorphic and splice variant forms of the protein wherein amino acid substitutions or deletions within the amino acid sequence do not affect the essential functioning of the protein. The term also includes functional fragments of the protein.

25 **Mutant *E5-1* gene** - the *E5-1* gene containing one or more mutations which lead to Alzheimer's Disease. This term is understood to include the various mutations that exist, wherein nucleotide substitutions in the gene sequence affect the essential function of the gene product.

30 **Mutant *E5-1* protein** - a protein analogous to *E5-1* protein but wherein one or more amino acid deletions and/or substitutions result in impairment of its essential function such that mammals, especially humans, whose *E5-1*-producing cells express mutant

E5-1 protein demonstrate the symptoms of Alzheimer's Disease.

Linkage analysis- Analysis of co-segregation of a disease trait or disease gene with polymorphic genetic markers of defined chromosomal location.

hARMP gene - human ARMP gene

5 **ORF** - open reading frame.

PCR - polymerase chain reaction.

contig - continuous cloned regions

YAC - yeast artificial chromosome

RT-PCR - reverse transcription polymerase chain reaction.

10 **SSR** - Simple sequence repeat polymorphism.

The present invention is concerned with the identification and sequencing of the mammalian ARMP gene in order to gain insight into the cause and etiology of familial Alzheimer's Disease. From this information, screening methods and therapies for the diagnosis and treatment of the disease can be developed. The gene
15 has been identified, cDNA isolated and cloned, and its transcripts and gene products identified and sequenced. During such identification of the gene, considerable sequence information has also been developed on intron information in the ARMP gene, flanking untranslated information and signal information and information involving neighbouring genes in the AD3 chromosome region. Direct sequencing of
20 overlapping RT-PCR products spanning the human gene isolated from affected members of large pedigrees linked to chromosome 14 has led to the discovery of missense mutations which co-segregate with the disease.

Although it is generally understood that Alzheimer's Disease is a neurological disorder, most likely in the brain, expression of ARMP has been found in varieties
25 of human tissue such as heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. Although this gene is expressed widely, the clinically apparent phenotype exists in brain although it is conceivable that biochemical phenotypes may exist in these other tissues. As with other genetic diseases such as Huntington's Disease and APP - Alzheimer's, the clinical disease manifestation may reflect
30 different biochemistries of different cell types and tissues (which stem from genetics

and the protein). Such findings suggest that AD may not be solely a neurological disorder but may also be a systemic disorder, hence requiring alternative therapeutic strategies which may be targeted to other tissues or organs or generally in addition or separately from neuronal or brain tissues.

5 The ARMP mutations identified have been related to Alzheimer disease pathology. With the identification of sequencing of the gene and the gene product, probes and antibodies raised to the gene product can be used in a variety of hybridization and immunological assays to screen for and detect the presence of either a normal or mutated gene or gene product.

10 Patient therapy through removal or blocking of the mutant gene product, as well as supplementation with the normal gene product by amplification, by genetic and recombinant techniques or by immunotherapy can now be achieved. Correction or modification of the defective gene product by protein treatment immunotherapy (using antibodies to the defective protein) or knock-out of the mutated gene is now
15 also possible. Familial Alzheimer's Disease could also be controlled by gene therapy in which the gene defect is corrected *in situ* or by the use of recombinant or other vehicles to deliver a DNA sequence capable of expressing the normal gene product, or a deliberately mutated version of the gene product whose effect counter balances the deleterious consequences of the disease mutation to the affected cells of the
20 patient.

 The present invention is also concerned with the identification and sequencing of a second gene, the *E5-1* gene on chromosome 1, which is associated with familial Alzheimer's Disease.

25 Disease mechanism insights and therapies analogous to those described above in relation to the ARMP gene will be available as a result of the identification and isolation of the *E5-1* gene.

Isolating the Human ARMP Gene

Genetic mapping of the AD3 locus.

30 After the initial regional mapping of the AD3 gene locus to 14q24.3 near the

anonymous microsatellite markers *D14S43* and *D14S53* (Schellenberg, GD et al., 1992; St George-Hyslop, P et al., 1992; Van Broeckhoven, C et al., 1992), twenty one pedigrees were used to segregate AD as a putative autosomal dominant trait (St George-Hyslop, P et al., 1992) and to investigate the segregation of 18 additional genetic markers from the 14q24.3 region which had been organized into a high density genetic linkage map (Figure 1b) (Weissenbach et al., 1992; Gyapay et al., 1994). Pairwise maximum likelihood analyses previously published confirmed substantial cumulative evidence for linkage between FAD and all of these markers (Table 1). However, much of the genetic data supporting linkage to these markers were derived from six large early onset pedigrees FAD1 (Nee et al., 1983), FAD2 (Frommelt et al., 1991), FAD3 (Goudsmit et al., 1981; Pollen, 1993), FAD4 (Foncin et al., 1985), TOR1.1 (Bergamini, 1991) and 603 (Pericak-Vance et al., 1988) each of which provide at least one anonymous genetic marker from 14q24.3 (St. George-Hyslop, P. et al 1992).

In order to more precisely define the location of the AD3 gene relative to the known locations of the genetic markers from 14q24.3, recombinational landmarks were sought by direct inspection of the raw haplotype data only from genotyped affected members of the six pedigrees showing definitive linkage to chromosome 14. This selective strategy in this particular instance necessarily discards data from the reconstructed genotypes of deceased affected members as well as from elderly asymptomatic members of the large pedigrees, and takes no account of the smaller pedigrees of uncertain linkage status. However, this strategy is very sound because it also avoids the acquisition of potentially misleading genotype data acquired either through errors in the reconstructed genotypes of deceased affected members arising from non-paternity or sampling errors or from the inclusion of unlinked pedigrees.

Upon inspection of the haplotype data for affected subjects, members of the six large pedigrees whose genotypes were directly determined revealed obligate recombinants at *D14S48* and *D14S53*, and at *D14S258* and *D14S63*. The single recombinant at *D14S53*, which depicts a telomeric boundary for the FAD region, occurred in the same AD affected subject of the FAD1 pedigree who had previously

been found to be recombinant at several other markers located telomeric to *D14S53* including *D14S48* (St George-Hyslop, P et al., 1992). Conversely, the single recombinant at *D14S258*, which marks a centromeric boundary of the FAD region, occurred in an affected member of the FAD3 pedigree who was also recombinant at several other markers centromeric to *D14S258* including *D14S63*. Both recombinant subjects had unequivocal evidence of Alzheimer's disease confirmed through standard clinical tests for the illness in other affected members of their families, and the genotype of both recombinant subjects was informative and co-segregating at multiple loci within the interval centromeric to *D14S53* and telomeric to *D14S258*.

When the haplotype analyses were enlarged to include the reconstructed genotypes of deceased affected members of the six large pedigrees as well as data from the remaining fifteen pedigrees with probabilities for linkage of less than 0.95, several additional recombinants were detected at one or more marker loci within the interval between *D14S53* and *D14S258*. Thus, one additional recombinant was detected in the reconstructed genotype of a deceased affected member of each of three of the larger FAD pedigrees (FAD1, FAD2 and other related families), and eight additional recombinants were detected in affected members of five smaller FAD pedigrees. However, while some of these recombinants might have correctly placed the AD3 gene within a more defined target region, we were forced to regard these potentially closer "internal recombinants" as unreliable not only for the reasons discussed earlier, but also because they provided mutually inconsistent locations for the AD3 gene within the *D14S53-D14S258* interval.

Construction of a physical contig spanning the AD3 region.

As an initial step toward cloning the AD3 gene a contig of overlapping genomic DNA fragments cloned into yeast artificial chromosome vectors, phage artificial chromosome vectors and cosmid vectors was constructed (Figure 1b). FISH mapping studies using cosmids derived from the YAC clones 932c7 and 964f5 suggested that the interval most likely to carry the AD3 gene was at least five megabases in size. Because the large size of this minimal co-segregating region

would make positional cloning strategies intractable, additional genetic pointers were sought which focused the search for the AD3 gene to one or more subregions within the interval flanked by *D14S53* and *D14S258*. Haplotype analyses at the markers between *D14S53* and *D14S258* failed to detect statistically significant evidence for linkage disequilibrium and/or allelic association between the FAD trait and alleles at any of these markers, irrespective of whether the analyses were restricted to those pedigrees with early onset forms of FAD, or were generalized to include all pedigrees. This result was not unexpected given the diverse ethnic origins of our pedigrees. However, when pedigrees of similar ethnic descent were collated, direct inspection of the haplotypes observed on the disease bearing chromosome segregating in different pedigrees of similar ethnic origin revealed two clusters of marker loci (Table 2). The first of these clusters located centromeric to *D14S77* (*D14S786*, *D14S277* and *D14S268*) and spanned the 0.95 Mb physical interval contained in YAC 78842 (depicted as region B in figure 1c). The second cluster was located telomeric to *D14S77* (*D14S43*, *D14S273*, and *D14S76*) and spanned the ~ 1Mb physical interval included within the overlapping YAC clones 964c2, 74163, 797d11 and part of 854f5 (depicted as region A in figure 1c). Identical alleles were observed in at least two pedigrees from the same ethnic origin (Table 2). As part the strategy, it was reasoned that the presence of shared alleles at one of these groups of physically clustered marker loci might reflect the co-inheritance of a small physical region surrounding the ARMP gene on the original founder chromosome in each ethnic population. Significantly, each of the shared extended haplotypes were rare in normal caucasian populations and allele sharing was not observed at other groups of markers spanning similar genetic intervals elsewhere on chromosome 14q24.3.

Transcription mapping and preliminary analysis of candidate genes

To isolate expressed sequences encoded within both critical intervals, a direct selection strategy was used involving immobilized, cloned, human genomic DNA as the hybridization target to recover transcribed sequences from primary complementary DNA pools derived from human brain mRNA (Rommens et al., 1993).

Approximately 900 putative cDNA fragments of size 100 to 600 base pairs were recovered from regions A and B in figure 1c. These fragments were hybridized to Southern blots containing genomic DNAs from each of the overlapping YAC clones and genomic DNAs from humans and other mammals. This identified a subset of 151 clones which showed evidence for evolutionary conservation and/or for a complex structure which suggested that they were derived from spliced mRNA. The clones within this subset were collated on the basis of physical map location, cross-hybridization and nucleotide sequence, and were used to screen conventional human brain cDNA libraries for longer cDNAs. At least 19 independent cDNA clones over 1 kb in length were isolated and then aligned into a partial transcription map of the AD3 region (Figure 1d). Only three of these transcripts corresponded to known characterized genes (cFOS, dihydrolipoamide succinyl transferase and latent transforming growth factor binding protein 2).

Recovery of Potential Candidate Genes

Each of the open reading frame portions of the candidate genes were recovered by RT-PCR from mRNA isolated from post-mortem brain tissue of normal control subjects and from either post-mortem brain tissue or cultured fibroblast cell lines of affected members of six pedigrees definitively linked to chromosome 14. The RT-PCR products were then screened for mutations using chemical cleavage and restriction endonuclease fingerprinting single-strand sequence conformational polymorphism methods (Saleeba and Cotton, 1993; Liu and Sommer, 1995), and by direct nucleotide sequencing. With one exception, all of the genes examined, although of interest, were not unique to affected subjects, and did not co-segregate with the disease. The single exception was the candidate gene represented by clone S182 which contained a series of nucleotide changes not observed in normal subjects, but which altered the predicted amino acid sequence in affected subjects. Although nucleotide sequence differences were also observed in some of the other genes, most were in the 3' untranslated regions and none were unique to AD-affected subjects.

The remaining sequences, a subset of which are mapped in Figure 1b together

with additional putative transcriptional sequences not identified in Figure 1c, are identified in the sequence listings as 14 through 43. The Sequence ID Nos: 14 to 43 represent neighbouring genes or fragments of neighbouring genes adjacent the hARMP gene or possibly additional coding fragments arising from alternative splicing of the hARMP. Sequence ID Nos: 44-125, and 149-159 represent neighboring genomic fragments containing both exon and intron information. Such sequences are useful for creating primers, for creating diagnostic tests, creating altered regulatory sequences and use of adjacent genomic sequences to create better animal models.

10 Characterization of the hARMP gene

Hybridization of the S182 clone to northern blots identified a transcript expressed widely in many areas of brain and peripheral tissues as a major 3.0 kb transcript and a minor transcript of 7.0 kb (Figures 4 and 5). Although the identity of the ~ 7.0 kb transcript is unclear, two observations suggest that the ~ 3.0 kb transcript represents an active product of the gene. Hybridization of the S182 clone to northern blots containing mRNA from a variety of murine tissues, including brain, identifies only a single transcript identical in size to the ~ 3.0 kb human transcript. All of the longer cDNA clones recovered to date (2.6-2.8 kb), which include both 5' and 3' UTRs and which account for the ~ 3.0 kb band on the northern blot, have mapped exclusively to the same physical region of chromosome 14. From these experiments the ~ 7.0 kb transcript could represent either a rare alternately spliced or polyadenylated isoform of the ~ 3.0 kb transcript or could represent another gene with homology to S182.

The nucleotide sequence of the major transcript was determined from the consensus of eleven independent longer cDNA clones and from 3 independent clones recovered by standard 5' rapid amplification of cDNA ends and bears no significant homology to other human genes. The cDNA of the sequenced transcript is provided in Sequence ID No: 1 and the predicted amino acid sequence is provided in Sequence ID No: 2. The cDNA sequence of another sequenced human clone is also provided as Sequence ID NO:132 and its predicted amino acid sequence is provided in SEQ

ID NO:133.

Analysis of the 5' end of multiple cDNA clones and RT-PCR products as well as corresponding genomic clones indicates that the 5' UTR is contained within at least two exons and that transcription either begins from two different start sites and/or that one of the early 5' untranslated exons is alternatively spliced (Table 6). The longest predicted open reading frame contains 467 amino acids with a small alternatively spliced exon of 4 amino acids at 25 codons from the putative start codon (Table 3). This putative start codon is the first in phase ATG located 63 bp downstream of a TGA stop codon and lacks a classical Kozak consensus sequences around the first two in-phase ATG sequences (Rogaer et al, in preparation). Like other genes lacking classical 'strong' start codons, the putative 5' UTR of the human transcripts are rich in GC.

Comparison of the nucleic acid and predicted amino acid sequences with available databases using the BLAST alignment paradigms revealed modest amino acid similarity with the *C. elegans* sperm integral membrane protein *SPE-4* ($p = 1.5e^{-25}$, 24-37% identity over three groups of at least fifty residues) and weaker similarity to portions of several other membrane spanning proteins including mammalian chromogranin A and alpha subunit of mammalian voltage dependent calcium channels (Altschul et al., 1990). This clearly established that they are not the same gene. The amino-acid sequence similarities across putative transmembrane domains may occasionally yield alignment that simply arises from the limited number of hydrophobic amino acids, but there is also extended sequence alignment between S182 protein and *SPE-4* at several hydrophilic domains. Both the putative S182 protein and *SPE-4* are predicted to be of comparable size (467 and 465 residues, respectively) and to contain at least seven transmembrane domains with a large acidic domain preceding the final predicted transmembrane domain. The S182 protein does have a longer predicted hydrophilic region at the N terminus.

Further investigation of the hARMP has revealed a host of sequence fragments which form the hARMP gene and include intron sequence information, 5' end untranslated sequence information and 3' end untranslated sequence information

(Table 6). Such sequence fragments are identified in Sequence ID Nos. 6 to 13.

Mutations in the S182 transcript

5 Direct sequencing of overlapping RT-PCR products spanning the 3.0 kb S182 transcript isolated from affected members of the six large pedigrees linked to chromosome 14 led to the discovery of eight missense mutations in each of the six pedigrees (Table 7, Figure 2). Each of these mutations co-segregated with the disease in the respective pedigrees [Figures 3(a)(b)(c)(d)(e)], and were absent from 142 unrelated neurologically normal subjects drawn from the same ethnic origins as the FAD pedigrees (284 unrelated chromosomes).

10 The location of the gene within the physical interval segregating with AD3 trait, the presence of eight different missense mutations which co-segregate with the disease trait in six pedigrees definitively linked to chromosome 14, and the absence of these mutations in 284 independent normal chromosomes cumulatively confirms that the hARMP gene is the AD3 locus. Further biologic support for this hypothesis arises both from the fact that the residues mutated in FAD kindreds are conserved in evolution (Table 3) and occur in domains of the protein which are also highly conserved, and from the fact that the S182 gene product is expressed at high levels in most regions of the brain including the most severely affected with AD.

15 The DNA sequence for the hARMP gene as cloned has been incorporated into a plasmid Bluescript. This stable vector has been deposited at ATCC under accession number _____ on April 28, 1995.

20 Several mutations in the hARMP gene have been identified which cause a severe type of familial Alzheimer's Disease. One, or a combination of these mutations may be responsible for this form of Alzheimer's Disease as well as several other neurological disorders. The mutations may be any form of nucleotide sequence alteration or substitution. Specific disease causing mutations in the form of nucleotide and/or amino acid substitutions have been located, although we anticipate additional mutations will be found in other families. Each of these nucleotide substitutions occurred within the putative ORF of the S182 transcript, and would be predicted to

25

30

change the encoded amino acid at the following positions, numbering from the first putative initiation codon. The mutations are listed in respect of their nucleotide locations in Sequence ID No: 1 and Sequence ID NO:132 (an additional human clone) and amino acid locations in Sequence ID No: 2 and Sequence ID NO:135 (the additional human clone).

5

i)	685, A→C	Met	146	Leu
ii)	737, A→G	His	163	Arg
iii)	986, C→A	Ala	246	Glu
iv)	1105, C→G	Leu	286	Val
v)	1478, G→A	Cys	410	Tyr
vi)	1027, C→T	Ala	260	Val
vii)	1102, C→T	Ala	285	Val
viii)	1422, C→G	Leu	392	Val

15

The Met146Leu, Ala246Glu and Cys410Tyr mutations have not been detected in the genomic DNA of affected members of the eight remaining small early onset autosomal dominant FAD pedigrees or six additional families in our collection which express late FAD onset. We predict that such mutations would not commonly occur in late onset FAD which has been excluded by genetic linkage studies from the more aggressive form of AD linked to chromosome 14q24.3 (St George-Hyslop, P et al., 1992; Schellenberg et al., 1993). The His163Arg mutation has been found in the genomic DNA of affected members of one additional FAD pedigree for which positive but significant statistical evidence for linkage to 14 becomes established. Age of onset of affected members was consistent with affected individuals from families linked to chromosome 14.

20

25

Mutations Ala260Val, Ala285Val, and Leu392Val all occur within the acidic hydrophilic loop between putative transmembrane 6 (TM6) and transmembrane (TM7) (Figure 6). Two of the mutations (A260V; A285V) and the L286V mutation are also located in the alternative spliced domain.

30

All eight of the mutations can be assayed by a variety of strategies (direct nucleotide sequencing, allele specific oligos, ligation polymerase chain reaction,

SSCP, RFLPs etc.) using RT-PCR products representing the mature mRNA/cDNA sequence or genomic DNA. Allele specific oligos were chosen for assaying the mutations. For the A260V and the A285V mutations, genomic DNA carrying the exon was amplified using the same PCR primers and methods as for the L286V mutation. PCR products were then denatured and slot blotted to duplicate nylon membranes using the slot blot protocol described for the C410T mutation.

Of all of the nucleotide substitutions co-segregated with the disease in their respective pedigrees (figures 3a to 3e), none were seen in asymptomatic family members aged more than two standard deviations beyond the mean age of onset, and none were present on 284 chromosomes from unrelated neurologically normal subjects drawn from comparable ethnic origins.

Identification of an Alternative Splice Form of the ARMP Gene Product

During sequencing studies of RT-PCR products for the ARMP gene recovered from a variety of tissues, it was discovered that some peripheral tissues (principally white blood cells) demonstrated two alternative splice forms of the ARMP gene. One form is identical to the (putatively 467 amino acid) isoform constitutatively expressed in all brain regions. The alternative splice form results from the exclusion of the segment of the cDNA between base pairs 1018 to 1116 inclusive, and results in a truncated isoform of the ARMP protein wherein the hydrophobic part of the hydrophilic acidically-charged loop immediately C-terminal to TM6 is removed. This alternatively spliced isoform therefore is characterized by preservation of the sequence N-terminal to and including the tyrosine at position 256, changing of the aspartate at 257 to alanine, and splicing on to the C-terminal part of the protein from and including tyrosine 291. Such splicing differences are often associated with important functional domains of the proteins. This argues that this hydrophilic loop (and consequently the N-terminal hydrophilic loop with similar amino acid charge) is/are active functional domains of the ARMP product and thus sites for therapeutic targeting.

ARMP Protein

With respect to DNA SEQ ID NO.1 and DNA SEQ ID NO: 132, analysis of the sequence of overlapping cDNA clones predicted an ORF protein of 467 amino acids when read from the first in phase ATG start codon and a molecular mass of approximately 52.6 kDa as later described, due to either polymorphisms in the protein or alternate splicing of the transcript, the molecular weight of the protein can vary due to possible substitutions or deletions of amino acids.

The analysis of predicted amino acid sequence using the Hopp and Woods algorithm suggested that the protein product is a multispanning integral membrane protein such as a receptor, a channel protein, or a structural membrane protein. The absence of recognizable signal peptide and the paucity of glycosylation sites are noteworthy, and the hydropathy profile suggests that the protein is less likely to be a soluble protein with a highly compact three-dimensional structure.

The protein may be a cellular protein with a highly compact three dimensional structure in which respect is may be similar to APOE which is also related to Alzheimer's Disease. In light of this putative functional role, it is proposed that this protein be labelled as the Alzheimer Related Membrane Protein (ARMP). The protein also contains a number of potential phosphorylation sites, one of which is the consensus site for MAPkinase which is also involved in the hyperphosphorylation of *tau* during the conversion of normal *tau* to neurofibrillary tangles. This consensus sequence may provide a common putative pathway linking this protein and other known biochemical aspects of Alzheimer's Disease and would represent a likely therapeutic target. Review of the protein structure reveals two sequences YTPF (residues 115-119) and STPE (residues 353 - 356) which represent the 5/T-P motif which is the MAP kinase consensus sequence. Several other phosphorylation sites exist with concensus sequences for Protein Kinase C activity. Because protein kinase C activity is associated with differences in the metabolism of APP which are relevant to Alzheimer's Disease, these sites on the ARMP protein and homologues are sites for therapeutic targetting.

The N-terminal is characterized by a highly hydrophilic acidic charged domain

with several potential phosphorylation domains, followed sequentially by a hydrophobic membrane spanning domain of 19 residues; a charged hydrophilic loop, then five additional hydrophobic membrane spanning domains interspersed with short (5-20 residue) hydrophilic domains; an additional larger acidic hydrophilic charged
5 loop, and then at least one and possibly two other hydrophobic potentially membrane spanning domains culminating in a polar domain at the C-terminus (Table 4 and Figure 6B). The presence of seven membrane spanning domains is characteristic of several classes of G-coupled receptor proteins but is also observed with other proteins including channel proteins.

10 Comparison of the nucleic acid and predicted amino acid sequences with available databases using the BLAST alignment paradigms revealed amino acid similarity with the *C. elegans* sperm integral membrane protein *spe-4* and a similarity to several other membrane spanning proteins including mammalian chromogranin A and the α -subunit of mammalian voltage dependent calcium channels.

15 The similarity between the putative products of the *spe-4* and ARMP genes implies that they may have similar activities. The *SPE-4* protein of *C. elegans* appears to be involved in the formation and stabilization of the fibrous body-membrane organelle (FBMO) complex during spermatogenesis. The FBMO is a specialized Golgi-derived organelle, consisting of a membrane bound vesicle attached
20 to and partly surrounding a complex of parallel protein fibers and may be involved in the transport and storage of soluble and membrane-bound polypeptides. Mutations in *spe-4* disrupt the FBMO complexes and arrest spermatogenesis. Therefore the physiologic function of *spe-4* may be either to stabilize interactions between integral membrane budding and fusion events, or to stabilize interactions between the
25 membrane and fibrillary proteins during the intracellular transport of the FBMO complex during spermatogenesis. Comparable functions could be envisaged for the ARMP. The ARMP could be involved either in the docking of other membrane-bound proteins such as β APP, or the axonal transport and fusion budding of membrane-bound vesicles during protein transport such as in the golgi apparatus or
30 endosome-lysosome system. If correct, then mutations might be expected to result

in aberrant transport and processing of β APP and/or abnormal interactions with cytoskeletal proteins such as the microtubule-associated protein *Tau*. Abnormalities in the intracellular and in the extracellular disposition of both β APP and *Tau* are in fact an integral part of the neuropathologic features of Alzheimer's Disease.

5 Although the location of the ARMP mutations in highly conserved residues within conserved domains of the putative proteins suggests that they are pathogenic, at least three of these mutations are conservative which is commensurate with the onset of disease in adult life. Because none of the mutations observed so far are deletions or nonsense mutations that would be expected to cause a loss of function, we cannot
10 predict whether these mutations will have a dominant gain-of-function effect and promote aberrant processing of β APP or a dominant loss-of-function effect causing arrest of normal β APP processing.

An alternative possibility is that the ARMP gene product may represent a receptor or channel protein. Mutations of such proteins have been causally related
15 to several other dominant neurological disorders in both vertebrate (eg. Malignant hyperthermia, hyperkalemic periodic paralysis in humans) and in invertebrate organisms (deg-1(d) mutants in *C.elegans*). Although the pathology of these other disorders does not resemble that of Alzheimer's Disease there is evidence for functional abnormalities in ion channels in Alzheimer's Disease. For example,
20 anomalies have been reported in the tetra-ethylammonium-sensitive 113pS potassium channel and in calcium homeostasis. Perturbations in transmembrane calcium fluxes might be especially relevant in view of the weak homology between S182 and the α -ID subunit of voltage-dependent calcium channels and the observations that increases in intracellular calcium in cultured cells can replicate some of the biochemical
25 features of Alzheimer's Disease such as alteration in the phosphorylation of Tau-microtubule-associated protein and increased production of A β peptides.

As mentioned purified normal ARMP protein is characterized by a molecular weight of 52.6kDa. The normal ARMP protein, substantially free of other proteins, is encoded by the aforementioned SEQ. ID No. 1 and SEQ ID NO:132. As will be
30 later discussed, the ARMP protein and fragments thereof may be made by a variety

of methods. Purified mutant ARMP protein is characterized by FAD - associated phenotype (necrotic death, apoptic death, granulovascular degeneration, neurofibrillary degeneration, abnormalities or changes in the metabolism of APP, and Ca^{2+} , K^{+} , and glucose, and mitochondrial function and energy metabolism neurotransmitter metabolism, all of which have been found to be abnormal in human brain, and/or peripheral tissue cells in subjects with Alzheimer's Disease) in a variety of cells. The mutant ARMP, free of other proteins, is encoded by the mutant DNA sequence.

10 Description of the *E5-1* gene, a homologue of the ARMP gene

A gene, *E5-1*, with substantial nucleotide and amino acid homology to the ARMP gene was identified by using the nucleotide sequence of the cDNA for ARMP to search data bases using the BLASTN paradigm of Altschul et al. 1990. Three expressed sequence tagged sites (ESTs) identified by accession numbers T03796, R14600, and R05907 were located which had substantial homology ($p < 1.0 \times 10^{-100}$, greater than 97% identity over at least 100 contiguous base pairs).

Oligonucleotide primers were produced from these sequences and used to generate PCR products by reverse transcriptase PCR (RT-PCR). These short RT-PCR products were partially sequenced to confirm their identity with the sequences within the data base and were then used as hybridization probes to screen full-length cDNA libraries. Several different cDNA's ranging in size from 1 Kb to 2.3 Kb were recovered from a cancer cell cDNA library (CaCo-2) and from a human brain cDNA library (*E5-1*, G1-1, cc54, cc32).

The nucleotide sequence of these clones confirmed that all were derivatives of the same transcript (designated *E5-1*).

The gene encoding the *E5-1* transcript mapped to human chromosome 1 using hybrid mapping panels and to two clusters of CEPH Mega YAC clones which have been placed upon a physical contig map (YAC clones 750g7, 921d12 mapped by FISH to 1q41; and YAC clone 787g12 which also contains an EST for the leukemia associated phosphoprotein (*LAP18*) gene which has been mapped to 1p36.1-p35) (data

not shown).

Hybridization of the *E5-1* cDNA clones to Northern Blots detected an ~2.3 kilobase mRNA band in many tissues including regions of the brain, as well as a ~2.6Kb mRNA band in muscle, cardiac muscle and pancreas (Figure 7).

5 In skeletal muscle, cardiac muscle and pancreas, the *E5-1* gene is expressed at relatively higher levels than in brain and as two different transcripts of ~2.3 Kb and ~2.6 Kb. Both of the *E5-1* transcripts have sizes clearly distinguishable from that of the 2.7 Kb ARMP transcript, and did not cross-hybridize with ARMP probes at high stringency. The cDNA sequence of the *E5-1* gene is identified as Sequence
10 ID No. 136.

The longest ORF within the *E5-1* cDNA consensus nucleotide sequence predicts a polypeptide containing 448 amino acids (numbering from the first in-phase ATG codon which was surrounded by a GCC-agg-GCT-ATG-c Kozak consensus sequence) (Sequence ID No. 137).

15 A comparison of the amino acid sequences of hARMP and *E5-1* homologue protein are shown in Table 8. Identical residues are indicated by vertical lines. The locations of mutations in the *E5-1* gene are indicated by downward pointing arrows. The locations of the mutations in the hARMP gene are indicated by upward pointing arrows. Putative TM domains are in open ended boxes. The alternatively spliced
20 exons are denoted by superscripted (*E5-1*) or subscripted (hARMP) "**".

BLASTP alignment analyses also detected significant homology with SPE-4 of *C. elegans* ($P = 3.5e-26$; identity = 20-63% over five domains of at least 22 residues), and weak homologies to brain sodium channels (alpha III subunit) and to the alpha subunit of voltage dependent calcium channels from a variety of species ($P = 0.02$; identities 20-28% over two or more domains each of at least 35 residues)
25 (Altschul, 1990). These alignments are similar to those described above for the ARMP gene. However, the most striking homology to the *E5-1* protein was found with the amino acid sequence predicted for ARMP. ARMP and *E5-1* proteins share 63% overall amino acid sequence identity, and several domains display virtually
30 complete identity (Table 8). Furthermore, all eight residues mutated in ARMP in

subjects with *AD3* are conserved in the E5-1 protein (Table 8). As would be expected, hydrophobicity analyses suggest that both proteins also share a similar structural organization.

5 The similarity was greatest in several domains of the protein corresponding to the intervals between transmembrane domain 1 (TM1) and TM6, and from TM7 to the C-terminus of the ARMP gene. The main difference from ARMP is a difference in the size and amino acid sequence of the acidically-charged hydrophilic loop in the position equivalent to the hydrophilic loop between transmembrane domains TM6 and TM7 in the ARMP protein and in the sequence of the N-terminal
10 hydrophilic domains.

Thus, both proteins are predicted to possess seven hydrophobic putative transmembrane domains, and both proteins bear large acidic hydrophilic domains at the N-terminus and between TM6 and TM7 (Figs. 6 and 8). A further similarity arose from analysis of RT-PCR products from brain and muscle RNA, which revealed
15 that nucleotides 1153-1250 of the E5-1 transcript are alternatively spliced. These nucleotides encode amino acids 263-296, which are located within the TM6-TM7 loop domain of the putative E5-1 protein, and which share 94% sequence identity with the alternatively spliced residues 257-290 in ARMP.

The most noticeable differences between the two predicted amino acid sequences occur in the amino acid sequence in the central portion of the TM6-TM7
20 hydrophilic loop (residues 304-374 of ARMP; 310 - 355 of E5-1), and in the N-terminal hydrophilic domain (Table 8). By analogy, this domain is also less highly conserved between the murine and human ARMP genes (identity = 47/60 residues), and shows no similarity with the equivalent region of SPE-4.

25 A splice variant of the E5-1 cDNA sequence identified as Sequence ID No. 136 has also been found in all tissues examined. This splice variant lacks the triplet GAA at nucleotide positions 1338-1340.

A further variant has been found in one normal individual whose E5-1 cDNA had C replacing T at nucleotide position 626, which does not change the amino acid
30 sequence.

Mutations of the *E5-1* gene associated with Alzheimer's Disease

The strong similarity between ARMP and the *E5-1* gene product raised the possibility that the *E5-1* gene might be the site of disease-causing mutations in some of a small number of early onset AD pedigrees in which genetic linkage studies have excluded chromosomes 14, 19 and 21. RT-PCR was used to isolate cDNAs corresponding to the *E5-1* transcript from lymphoblasts, fibroblasts or post-mortem brain tissue of affected members of eight pedigrees with early onset familial AD (FAD) in which mutations in the β APP and ARMP gene had previously been excluded by direct sequencing studies.

Examination of these RT-PCR products detected a heterozygous A→G substitution at nucleotide 1080 in all four affected members of an extended pedigree of Italian origin (Flo10) with early onset, pathologically confirmed FAD (onset = 50-70 yrs). This mutation would be predicted to cause a Met→Val missense mutation at codon 239 (Table 8).

A second mutation (A→T at nucleotide 787) causing a Asn→Ile substitution at codon 141 was found in affected members of a group of related pedigrees of Volga German ancestry (represented by cell lines AG09369, AG09907, AG09952, and AG09905, Coriell Institute, Camden NJ). Significantly, one subject (AG09907) was homozygous for this mutation, an observation compatible with the in-bred nature of these pedigrees. Significantly, this subject did not have a significantly different clinical picture from those subjects heterozygous for the Arg141Ile mutation. Neither of the *E5-1* gene mutations were found in 284 normal Caucasian controls nor were they present in affected members of pedigrees with the *AD3* type of AD.

Both of these mutations would be predicted to cause substitution of residues which are highly conserved within the ARMP/*E5-1* gene family.

The finding of a gene whose product is predicted to share substantial amino acid and structural similarities with the ARMP gene product suggests that these proteins may be functionally related either as independent proteins with overlapping functions but perhaps with slightly different specific activities, as physically associated subunits of a multimeric polypeptide or as independent proteins performing

consecutive functions in the same pathway.

The observation of two different missense mutations in conserved domains of the E5-1 protein in subjects with a familial form of AD argues that these mutations are, like those in the ARMP gene, causal to AD. This conclusion is significant because, while the disease phenotypes associated with mutations in the ARMP gene (onset 30-50yrs, duration 10 years) are subtly different from that associated with mutations in the *E5-1* gene (onset 40-70 years; duration up to 20yrs), the general similarities clearly argue that the biochemical pathway subsumed by members of this gene family is central to the genesis of at least early onset AD. The subtle differences in disease phenotype may reflect a lower level of expression of the *E5-1* transcript in the CNS, or may reflect a different role for the *E5-1* gene product.

By analogy to the effects of ARMP mutations, E5-1 when mutated may cause aberrant processing of APP (Amyloid Precursor Protein) into A β peptide, hyperphosphorylation of Tau microtubule associated protein and abnormalities of intracellular calcium homeostasis. Interference with these anomalous interactions provides a potential therapy for AD.

Functional Domains of the ARMP Protein are Defined by Splicing Sites and Similarities within Other Members of a Gene Family

The ARMP protein is a member of a novel class of transmembrane proteins which share substantial amino acid homology. The homology is sufficient that certain nucleotide probes and antibodies raised against one can identify other members of this gene family. The major difference between members of this family reside in the amino acid and nucleotide sequence homologous to the hydrophilic acid loop domain between putative transmembrane 6 and transmembrane 7 domains of the ARMP gene and gene product. This region is alternatively spliced in some non-neural tissues, and is also the site of several pathogenic disease-causing mutations in the ARMP gene. The variable splicing of this hydrophilic loop, the presence of a high-density of pathogenic mutations within this loop, and the fact that the amino acid sequences of the loop differs between members of the gene family suggest that this loop is an

important functional domain of the protein and may confer some specificity to the physiologic and pathogenic interactions which the ARMP gene product undergoes because the N-terminal hydrophilic domain shares the same acidic charge and same orientation with respect to the membrane, it is very likely that these two domains share functionality either in a coordinated (together) or independent fashion (eg. different ligands or functional properties). As a result everything said about the hydrophilic loop shall apply also to the N-terminal hydrophilic domain.

Knowledge of the specificity of the loop can be used to identify ligands and functional properties of the ARMP gene product (eg. sites of interactions with APP, cytosolic proteins such as kinases, *Tau*, and MAP, etc.). Soluble recombinant fusion proteins can be made or the nucleotide sequence coding for amino acids within the loop or parts of the loop can be expressed in suitable vectors (yeast-2-hybrid, baculovirus, and phage - display systems for instance), and used to identify other proteins which interact with ARMP in the pathogenesis of Alzheimer's disease and other neurological and psychiatric diseases. Therapies can be designed to modulate these interactions and thus to modulate Alzheimer's disease and the other conditions associated with acquired or inherited abnormalities of the ARMP gene or its gene products. The potential efficacy of these therapies can be tested by analyzing the affinity and function of these interactions after exposure to the therapeutic agent by standard pharmacokinetic measurements of affinity (K_d and V_{max} etc) using synthetic peptides or recombinant proteins corresponding to functional domains of the ARMP gene (or its homologues). An alternate method for assaying the effect of any interactions involving functional domains such as the hydrophilic loop is to monitor changes in the intracellular trafficking and post-translational modification of the ARMP gene by *in-situ* hybridization, immunohistochemistry, Western blotting and metabolic pulse-chase labelling studies in the presence of and in the absence of the therapeutic agents. A third way is to monitor the effects of "downstream" events including (i) changes in the intracellular metabolism, trafficking and targeting of APP and its products; (ii) changes in second messenger event eg. cAMP, intracellular Ca^{++} protein kinase activities, etc..

Isolation and Purification of the ARMP Protein

5 The ARMP protein may be isolated and purified by methods selected on the basis of properties revealed by its sequence. Since the protein possesses properties of a membrane-spanning protein, a membrane fraction of cells in which the protein is highly expressed (eg. central nervous system cells or cells from other tissues) would be isolated and the proteins removed by extraction and the proteins solubilized using a detergent.

10 Purification can be achieved using protein purification procedures such as chromatography methods (gel-filtration, ion-exchange and immunoaffinity), by high-performance liquid chromatography (RP-HPLC, ion-exchange HPLC, size-exclusion HPLC, high-performance chromatofocusing and hydrophobic interaction chromatography) or by precipitation (immunoprecipitation). Polyacrylamide gel electrophoresis can also be used to isolate the ARMP protein based on its molecular weight, charge properties and hydrophobicity.

15 Similar procedures to those just mentioned could be used to purify the protein from cells transfected with vectors containing the ARMP gene (eg. baculovirus systems, yeast expression systems, eukaryotic expression systems).

20 Purified protein can be used in further biochemical analyses to establish secondary and tertiary structure which may aid in the design of pharmaceuticals to interact with the protein, alter protein charge configuration or charge interaction with other proteins, lipid or saccharide moieties, alter its function in membranes as a transporter channel or receptor and/or in cells as an enzyme or structural protein and treat the disease.

25 The protein can also be purified by creating a fusion protein by legating the ARMP cDNA sequence to a vector which contains sequence for another peptide (eg. GST - glutathionine succinyl transferase). The fusion protein is expressed and recovered from prokaryotic (eg. bacterial or baculovirus) or eukaryotic cells. The fusion protein can then be purified by affinity chromatography based upon the fusion vector sequence. The ARMP protein can then be further purified from the fusion

30

protein by enzymatic cleavage of the fusion protein.

Isolating mouse ARMP gene

5 In order to characterize the physiological significance of the normal and mutant hARMP gene and gene products in a transgenic mouse model it was necessary to recover a mouse homologue of the hARMP gene. We recovered a murine homologue for the hARMP gene by screening a mouse cDNA library with a labelled human DNA probe and in this manner recovered a 2 kb partial transcript (representing the 3' end of the gene) and several RT-PCR products representing the
10 5'end. Sequencing of the consensus cDNA transcript of the murine homologue revealed substantial amino acid identity. The sequence cDNA is identified in Sequence ID No. 3 and the predicted amino acid Sequence is provided in Sequence ID No. 4. Further sequencing of the mouse cDNA transcript has provided the sequence for the complete coding sequence identified as SEQ ID NO:134 and the
15 predicted amino acid sequence from this sequence is provided in SEQ ID NO: 135. More importantly, all of the amino acids that were mutated in the FAD pedigrees were conserved between the murine homologue and the normal human variant (Table 3). This conservation of the ARMP gene as is shown in table 3, indicates that an orthologous gene exists in the mouse (mARMP), and it is now possible to clone
20 mouse genomic libraries using human ARMP probes. This will also make it possible to identify and characterize the ARMP gene in other species. This also provides evidence of animals with various disease states or disorders currently known or yet to be elucidated.

25

Transgenic Mouse Model

The creation of a mouse model for Alzheimer's Disease is important to the understanding of the disease and for the testing of possible therapies. Currently no unambiguous viable animal model for Alzheimer's Disease exists.

30

There are several ways in which to create an animal model for Alzheimer's

Disease. Generation of a specific mutation in the mouse gene such as the identified hARMP gene mutations is one strategy. Secondly, we could insert a wild type human gene and/or humanize the murine gene by homologous recombination. Thirdly, it is also possible to insert a mutant (single or multiple) human gene as genomic or minigene cDNA constructs using wild type or mutant or artificial promoter elements. Fourthly, knock-out of the endogenous murine genes may be accomplished by the insertion of artificially modified fragments of the endogenous gene by homologous recombination. The modifications include insertion of mutant stop codons, the deletion of DNA sequences, or the inclusion of recombination elements (lox p sites) recognized by enzymes such as Cre recombinase.

To inactivate the mARMP gene chemical or x-ray mutagenesis of mouse gametes, followed by fertilization, can be applied. Heterozygous offspring can then be identified by Southern blotting to demonstrate loss of one allele by dosage, or failure to inherit one parental allele using RFLP markers.

To create a transgenic mouse a mutant version of hARMP or mARMP can be inserted into a mouse germ line using standard techniques of oocyte microinjection or transfection or microinjection into stem cells. Alternatively, if it is desired to inactivate or replace the endogenous mARMP gene, homologous recombination using embryonic stem cells may be applied.

For oocyte injection, one or more copies of the mutant or wild type ARMP gene can be inserted into the pronucleus of a just-fertilized mouse oocyte. This oocyte is then reimplanted into a pseudo-pregnant foster mother. The liveborn mice can then be screened for integrants using analysis of tail DNA for the presence of human ARMP gene sequences. The transgene can be either a complete genomic sequence injected as a YAC, BAC, PAC or other chromosome DNA fragment, a cDNA with either the natural promoter or a heterologous promoter, or a minigene containing all of the coding region and other elements found to be necessary for optimum expression.

Retroviral infection of early embryos can also be done to insert the mutant or wild type hARMP. In this method, the mutant or wild type hARMP is inserted into

a retroviral vector which is used to directly infect mouse embryos during the early stages of development to generate a chimera, some of which will lead to germline transmission. Similar experiments can be conducted in the case of mutant proteins, using mutant murine or other animal ARMP gene sequences.

5 Homologous recombination using stem cells allows for the screening of gene transfer cells to identify the rare homologous recombination events. Once identified, these can be used to generate chimeras by injection of mouse blastocysts, and a proportion of the resulting mice will show germline transmission from the recombinant line. This methodology is especially useful if inactivation of the
10 mARMP gene is desired. For example, inactivation of the mARMP gene can be done by designing a DNA fragment which contains sequences from a mARMP exon flanking a selectable marker. Homologous recombination leads to the insertion of the marker sequences in the middle of an exon, inactivating the mARMP gene. DNA analysis of individual clones can then be used to recognize the homologous
15 recombination events.

It is also possible to create mutations in the mouse germline by injecting oligonucleotides containing the mutation of interest and screening the resulting cells by PCR.

This embodiment of the invention has the most significant commercial value
20 as a mouse model for Alzheimer's Disease. Because of the high percentage of sequence conservation between human and mouse it is contemplated that an orthologous gene will exist also in many other species. It is thus contemplated that it will be possible to generate other animal models using similar technology.

25 Screening and Diagnosis for Alzheimer's Disease
General Diagnostic Uses of the ARMP Gene and Gene Product

The ARMP gene and gene products will be useful for diagnosis of Alzheimer's disease, presenile and senile dementias, psychiatric diseases such as schizophrenia, depression, etc., and neurologic diseases such as stroke and cerebral hemorrhage -
30 all of which are seen to a greater or lesser extent in symptomatic subjects bearing

mutations in the ARMP gene or in the APP gene. Diagnosis of inherited cases of these diseases can be accomplished by analysis of the nucleotide sequence (including genomic and cDNA sequences included in this patent). Diagnosis can also be achieved by monitoring alterations in the electrophoretic mobility and by the reaction with specific antibodies to mutant or wild-type ARMP gene products, and by functional assays demonstrating altered function of the ARMP gene product. In addition, the ARMP gene and ARMP gene products can be used to search for inherited anomalies in the gene and/or its products (as well as those of the homologous gene) and can also be used for diagnosis in the same way as they can be used for diagnosis of non-genetic cases.

Diagnosis of non-inherited cases can be made by observation of alterations in the ARMP transcription, translation, and post-translational modification and processing as well as alterations in the intracellular and extracellular trafficking of ARMP gene products in the brain and peripheral cells. Such changes will include alterations in the amount of ARMP messenger RNA and/or protein, alteration in phosphorylation state, abnormal intracellular location/distribution, abnormal extracellular distribution, etc. Such assays will include: Northern Blots (with ARMP-specific and ARMP-non-specific nucleotide probes which also cross-react with other members of the gene family), and Western blots and enzyme-linked immunosorbent assays (ELISA) (with antibodies raised specifically to: ARMP; to various functional domains of ARMP; to other members of the homologous gene family; and to various post-translational modification states including glycosylated and phosphorylated isoforms). These assays can be performed on peripheral tissues (eg. blood cells, plasma, cultured or other fibroblast tissues, etc.) as well as on biopsies of CNS tissues obtained antemortem or postmortem, and upon cerebrospinal fluid. Such assays might also include in-situ hybridization and immunohistochemistry (to localized messenger RNA and protein to specific subcellular compartments and/or within neuropathological structures associated with these diseases such as neurofibrillary tangles and amyloid plaques).

Screening for Alzheimer's Disease

Screening for Alzheimer's Disease as linked to chromosome 14 may now be readily carried out because of the knowledge of the mutations in the gene.

5 People with a high risk for Alzheimer's Disease (present in family pedigree) or, individuals not previously known to be at risk, or people in general may be screened routinely using probes to detect the presence of a mutant ARMP gene by a variety of techniques. Genomic DNA used for the diagnosis may be obtained from body cells, such as those present in the blood, tissue biopsy, surgical specimen, or autopsy material. The DNA may be isolated and used directly for detection of a
10 specific sequence or may be PCR amplified prior to analysis. RNA or cDNA may also be used. To detect a specific DNA sequence hybridization using specific oligonucleotides, direct DNA sequencing, restriction enzyme digest, RNase protection, chemical cleavage, and ligase-mediated detection are all methods which can be utilized. Oligonucleotides specific to mutant sequences can be chemically
15 synthesized and labelled radioactively with isotopes, or non-radioactively using biotin tags, and hybridized to individual DNA samples immobilized on membranes or other solid-supports by dot-blot or transfer from gels after electrophoresis. The presence or absence of these mutant sequences are then visualized using methods such as autoradiography, fluorometry, or colorimetric reaction. Examples of suitable PCR
20 primers which are useful for example in amplifying portions of the subject sequence containing the aforementioned mutations are set out in Table 5. This table also sets out the change in enzyme site to provide a useful diagnostic tool as defined herein.

Direct DNA sequencing reveals sequence differences between normal and mutant ARMP DNA. Cloned DNA segments may be used as probes to detect
25 specific DNA segments. PCR can be used to enhance the sensitivity of this method. PCR is an enzymatic amplification directed by sequence-specific primers, and involves repeated cycles of heat denaturation of the DNA, annealing of the complementary primers and extension of the annealed primer with a DNA polymerase. This results in an exponential increase of the target DNA.

30 Other nucleotide sequence amplification techniques may be used, such as

ligation-mediated PCR, anchored PCR and enzymatic amplification as would be understood by those skilled in the art.

5 Sequence alterations may also generate fortuitous restriction enzyme recognition sites which are revealed by the use of appropriate enzyme digestion followed by gel-blot hybridization. DNA fragments carrying the site (normal or mutant) are detected by their increase or reduction in size, or by the increase or decrease of corresponding restriction fragment numbers. Genomic DNA samples may also be amplified by PCR prior to treatment with the appropriate restriction enzyme and the fragments of different sizes are visualized under UV light in the presence of
10 ethidium bromide after gel electrophoresis.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. Small deletions may also be detected as changes in the migration
15 pattern of DNA heteroduplexes in non-denaturing gel electrophoresis. Alternatively, a single base substitution mutation may be detected based on differential PCR product length in PCR. The PCR products of the normal and mutant gene could be differentially detected in acrylamide gels.

20 Nuclease protection assays (S1 or ligase-mediated) also reveal sequence changes at specific locations.

Alternatively, to confirm or detect a polymorphism restriction mapping changes ligated PCR, ASO, REF-SSCP chemical cleavage, endonuclease cleavage at mismatch sites and SSCP may be used. Both REF-SSCP and SSCP are mobility shift assays which are based upon the change in conformation due to mutations.

25 DNA fragments may also be visualized by methods in which the individual DNA samples are not immobilized on membranes. The probe and target sequences may be in solution or the probe sequence may be immobilized. Autoradiography, radioactive decay, spectrophotometry, and fluorometry may also be used to identify specific individual genotypes. Finally, mutations can be detected by direct nucleotide
30 sequencing.

According to an embodiment of the invention, the portion of the cDNA or genomic DNA segment that is informative for a mutation, can be amplified using PCR. For example, the DNA segment immediately surrounding the C 410 Y mutation acquired from peripheral blood samples from an individual can be screened using the oligonucleotide primers 885 (tgagactggaacacaac) sequence ID No:128 and 893 (gtgtggccagggtagagaact) sequence ID No:129. This region would then be amplified by PCR, the products separated by electrophoresis, and transferred to membrane. Labelled oligonucleotide probes are then hybridized to the DNA fragments and autoradiography performed.

ARMP Expression

As an embodiment of the present invention, ARMP protein may be expressed using eukaryotic and prokaryotic expression systems. Eukaryotic expression systems can be used for many studies of the ARMP gene and gene product including determination of proper expression and post-translational modifications for full biological activity, identifying regulatory elements located in the 5' region of the ARMP gene and their role in tissue regulation of protein expression, production of large amounts of the normal and mutant protein for isolation and purification, to use cells expressing the ARMP protein as a functional assay system for antibodies generated against the protein or to test effectiveness of pharmacological agents, or as a component of a signal transduction system, to study the function of the normal complete protein, specific portions of the protein, or of naturally occurring and artificially produced mutant proteins.

Eukaryotic and prokaryotic expression systems were generated using two different classes of ARMP nucleotide cDNA sequence inserts. In the first class, termed full-length constructs, the entire ARMP cDNA sequence is inserted into the expression plasmid in the correct orientation, and includes both the natural 5' UTR and 3' UTR sequences as well as the entire open reading frame. The open reading frames bear a nucleotide sequence cassette which allows either the wild type open reading frame to be included in the expression system or alternatively, single or a

combination of double mutations can be inserted into the open reading frame. This was accomplished by removing a restriction fragment from the wild type open reading frame using the enzymes *NarI* and *PflmI* and replacing it with a similar fragment generated by reverse transcriptase PCR and which bears the nucleotide sequence encoding either the Met146Leu mutation or the Hys163Arg mutation. A second restriction fragment was removed from the wild type normal nucleotide sequence for the open reading frame by cleavage with the enzymes *PflmI* and *NcoI* and replaced with restriction fragments bearing either the nucleotide sequence encoding the Ala246Glu mutation, or the Ala260Val mutation or the Ala285Val mutation or the Leu286Val mutation, or the Leu392Val mutation, or the Cys410Tyr mutation. Finally, a third variant bearing combinations of either the Met146Leu or His163Arg mutations in tandem with the remaining mutations, was made by linking the *NarI*-*PflmI* fragment bearing these mutations and the *PflmI*-*NcoI* fragments bearing the remaining mutations.

A second variant of cDNA inserts bearing wild type or mutant cDNA sequences was constructed by removing from the full-length cDNA the 5' UTR and part of the 3' UTR sequences. The 5' UTR sequence was replaced with a synthetic oligonucleotide containing a *KpnI* restriction site and a Kozak initiation site (oligonucleotide 969: ggtaccgccaccatgacagaggtacctgcac, Sequence ID No:138). The 3' UTR was replaced with an oligonucleotide corresponding to position 2566 of the cDNA and bears an artificial *EcoRI* site (oligonucleotide 970:gaattcactggctgtagaaaaagac, Sequence ID No:139). Mutant variants of this construct were then made by inserting the same mutant sequences described above at the *NarI*-*PflmI* fragment, and at the *PflmI*-*NcoI* sites described above.

For eukaryotic expressions, these various cDNA constructs bearing wild type and mutant sequences described above were cloned into the expression vector pZeoSV (invitrogen). For prokaryotic expression, two constructs have been made using the glutathione S-transferase fusion vector pGEX-kg. The inserts which have been attached to the GST fusion nucleotide sequence are the same nucleotide sequence described above (generated with the oligonucleotide primers 969, Sequence ID

No:138 and 970, Sequence ID No: 139) bearing either the normal open reading frame nucleotide sequence, or bearing a combination of single and double mutations as described above. This construct allows expression of the full-length protein in mutant and wild type variants in prokaryotic cell systems as a GST fusion protein which allows purification of the full-length protein followed by removal of the GST fusion product by thrombin digestion. The second prokaryotic cDNA construct was generated to create a fusion protein with the same vector, and allows the production of the amino acid sequence corresponding to the hydrophillic acidic loop domain between TM6 and TM7 of the full-length protein, as either a wild type nucleotide sequence (thus a wild type amino acid sequence for fusion proteins) or as a mutant sequence bearing either the Ala285Val mutation, or the Leu286Val mutation, or the Leu392Val mutation. This was accomplished by recovering wild type or mutant sequence from appropriate sources of RNA using the oligonucleotide primers 989:ggatccgggtccacttcgtatgctg, Sequence ID No:140, and 990:tttttgaattcttaggctatggtgtgtcca, Sequence ID No: 141. This allows cloning of the appropriate mutant or wild type nucleotide sequence corresponding to the hydrophillic acid loop domain at the BamHI and the EcoRI sites within the pGEX-KG vector.

These prokaryotic expression systems allow the holo-protein or various important functional domains of the protein to be recovered as fusion proteins and then used for binding studies, structural studies, functional studies, and for the generation of appropriate antibodies.

Expression of the ARMP gene in heterologous cell systems can be used to demonstrate structure-function relationships. Ligating the ARMP DNA sequence into a plasmid expression vector to transfect cells is a useful method to test the proteins influence on various cellular biochemical parameters. Plasmid expression vectors containing either the entire, normal or mutant human or mouse ARMP sequence or portions thereof, can be used in *in vitro* mutagenesis experiments which will identify portions of the protein crucial for regulatory function.

The DNA sequence can be manipulated in studies to understand the expression of the gene and its product, to achieve production of large quantities of the protein

for functional analysis, for antibody production, and for patient therapy. The changes in the sequence may or may not alter the expression pattern in terms of relative quantities, tissue-specificity and functional properties. Partial or full-length DNA sequences which encode for the ARMP protein, modified or unmodified, may be
5 ligated to bacterial expression vectors. *E. coli* can be used using a variety of expression vector systems, eg. the T7 RNA polymerase/promoter system using two plasmids or by labeling of plasmid-encoded proteins, or by expression by infection with M13 Phage mGPI-2. *E. coli* vectors can also be used with Phage lambda regulatory sequences, by fusion protein vectors (eg. lacZ and trpE), by maltose-
10 binding protein fusions, and by glutathione-S-transferase fusion proteins, etc., all of which together with many other prokaryotic expression systems are widely available commercially.

Alternatively, the ARMP protein can be expressed in insect cells using baculoviral vectors, or in mammalian cells using vaccinia virus or specialised
15 eukaryotic expression vectors. For expression in mammalian cells, the cDNA sequence may be ligated to heterologous promoters, such as the simian virus (SV40) promoter in the pSV2 vector and other similar vectors and introduced into cultured eukaryotic cells, such as COS cells to achieve transient or long-term expression. The stable integration of the chimeric gene construct may be maintained in mammalian
20 cells by biochemical selection, such as neomycin and mycophenolic acid.

The ARMP DNA sequence can be altered using procedures such as restriction enzyme digestion, fill-in with DNA polymerase, deletion by exonuclease, extension by terminal deoxynucleotide transferase, ligation of synthetic or cloned DNA sequences and site-directed sequence alteration with the use of specific
25 oligonucleotides together with PCR.

The cDNA sequence or portions thereof, or a mini gene consisting of a cDNA with an intron and its own promoter, is introduced into eukaryotic expression vectors by conventional techniques. These vectors permit the transcription of the cDNA in eukaryotic cells by providing regulatory sequences that initiate and enhance the
30 transcription of the cDNA and ensure its proper splicing and polyadenylation. The

endogenous ARMP gene promoter can also be used. Different promoters within vectors have different activities which alters the level of expression of the cDNA. In addition, certain promoters can also modulate function such as the glucocorticoid-responsive promoter from the mouse mammary tumor virus.

5 Some of the vectors listed contain selectable markers or neo bacterial genes that permit isolation of cells by chemical selection. Stable long-term vectors can be maintained in cells as episomal, freely replicating entities by using regulatory elements of viruses. Cell lines can also be produced which have integrated the vector into the genomic DNA. In this manner, the gene product is produced on a continuous
10 basis.

Vectors are introduced into recipient cells by various methods including calcium phosphate, strontium phosphate, electroporation, lipofection, DEAE dextran, microinjection, or by protoplast fusion. Alternatively, the cDNA can be introduced by infection using viral vectors.

15 Using the techniques mentioned, the expression vectors containing the ARMP gene or portions thereof can be introduced into a variety of mammalian cells from other species or into non-mammalian cells.

The recombinant cloning vector, according to this invention, comprises the selected DNA of the DNA sequences of this invention for expression in a suitable
20 host. The DNA is operatively linked in the vector to an expression control sequence in the recombinant DNA molecule so that normal and mutant ARMP protein can be expressed. The expression control sequence may be selected from the group consisting of sequences that control the expression of genes of prokaryotic or eukaryotic cells and their viruses and combinations thereof. The expression control
25 sequence may be selected from the group consisting of the lac system, the trp system, the tac system, the trc system, major operator and promoter regions of phage lambda, the control region of the fd coat protein, early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus, simian virus, 3-phosphoglycerate kinase promoter, yeast acid phosphatase promoters, yeast alpha-mating factors and combinations thereof.
30

The host cell which may be transfected with the vector of this invention may be selected from the group consisting of *E.coli*, *pseudomonas*, *bacillus subtilis*, *bacillus stearothermophilus*, or other bacilli; other bacteria, yeast, fungi, insect, mouse or other animal, plant hosts, or human tissue cells.

- 5 For the mutant ARMP DNA sequence similar systems are employed to express and the produce the mutant protein.

Antibodies to Detect ARMP

- 10 Antibodies to epitopes with the ARMP protein can be raised to provide information on the characteristics of the proteins. Generation of antibodies would enable the visualization of the protein in cells and tissues using Western blotting. In this technique, proteins are run on polyacrylamide gel and then transferred onto nitrocellulose membranes. These membranes are then incubated in the presence of the antibody (primary), then following washing are incubated to a secondary antibody
15 which is used for detection of the protein-primary antibody complex. Following repeated washing, the entire complex is visualized using colourimetric or chemiluminescent methods.

- 20 Antibodies to the ARMP protein also allow for the use of immunocytochemistry and immunofluorescence techniques in which the proteins can be visualized directly in cells and tissues. This is most helpful in order to establish the subcellular location of the protein and the tissue specificity of the protein.

- 25 In order to prepare polyclonal antibodies, fusion proteins containing defined portions or all of the ARMP protein can be synthesized in bacteria by expression of corresponding DNA sequences in a suitable cloning vehicle. The protein can then be purified, coupled to a carrier protein and mixed with Freund's adjuvant (to help stimulate the antigenic response by the rabbits) and injected into rabbits or other laboratory animals. Alternatively, protein can be isolated from cultured cells expressing the protein. Following booster injections at bi-weekly intervals, the rabbits or other laboratory animals are then bled and the sera isolated. The sera can
30 be used directly or purified prior to use, by various methods including affinity

chromatography, Protein A-Sepharose, Antigen Sepharose, Anti-mouse-Ig-Sepharose. The sera can then be used to probe protein extracts run on a polyacrylamide gel to identify the ARMP protein. Alternatively, synthetic peptides can be made to the antigenic portions of the protein and used to inoculate the animals.

5 To produce monoclonal ARMP antibodies, cells actively expressing the protein are cultured or isolated from tissues and the cell membranes isolated. The membranes, extracts, or recombinant protein extracts, containing the ARMP protein, are injected in Freund's adjuvant into mice. After being injected 9 times over a three
10 week period, the mice spleens are removed and resuspended in phosphate buffered saline (PBS). The spleen cells serve as a source of lymphocytes, some of which are producing antibody of the appropriate specificity. These are then fused with a permanently growing myeloma partner cell, and the products of the fusion are plated into a number of tissue culture wells in the presence of a selective agent such as HAT. The wells are then screened to identify those containing cells making useful
15 antibody by ELISA. These are then freshly plated. After a period of growth, these wells are again screened to identify antibody-producing cells. Several cloning procedures are carried out until over 90% of the wells contain single clones which are positive for antibody production. From this procedure a stable line of clones is established which produce the antibody. The monoclonal antibody can then be
20 purified by affinity chromatography using Protein A Sepharose, ion-exchange chromatography, as well as variations and combinations of these techniques.

In situ hybridization is another method used to detect the expression of ARMP protein. *In situ* hybridization relies upon the hybridization of a specifically labelled nucleic acid probe to the cellular RNA in individual cells or tissues. Therefore, it
25 allows the identification of mRNA within intact tissues, such as the brain. In this method, oligonucleotides corresponding to unique portions of the ARMP gene are used to detect specific mRNA species in the brain.

 In this method a rat is anesthetized and transcardially perfused with cold PBS, followed by perfusion with a formaldehyde solution. The brain or other tissues is
30 then removed, frozen in liquid nitrogen, and cut into thin micron sections. The

sections are placed on slides and incubated in proteinase K. Following rinsing in DEP, water and ethanol, the slides are placed in prehybridization buffer. A radioactive probe corresponding to the primer is made by nick translation and incubated with the sectioned brain tissue. After incubation and air drying, the labeled areas are visualized by autoradiography. Dark spots on the tissue sample indicate hybridization of the probe with brain mRNA which demonstrates the expression of the protein.

Antibodies may also be used coupled to compounds for diagnostic and/or therapeutic uses such as radionuclides for imaging and therapy and liposomes for the targeting of compounds to a specific tissue location.

Isolation and Purification of E5-1 protein

The E5-1 protein may be isolated and purified by the types of methods described above for the ARMP protein.

The protein may also be prepared by expression of the E5-1 cDNA described herein in a suitable host. The protein is preferably expressed as a fusion protein by ligating its encoding cDNA sequence to a vector containing the coding sequence for another suitable peptide, eg. GST. The fusion protein is expressed and recovered from prokaryotic cells such as bacterial or baculovirus cells or from eukaryotic cells. Antibodies to ARMP, by virtue of portions of amino acid sequence identity with E5-1, can be used to purify, attract and bind to E5-1 protein and vice versa.

Transgenic Mouse Model of E5-1 related Alzheimer's Disease

An animal model of Alzheimer's disease related to mutations of the E5-1 gene may be created by methods analogous to those described above for the ARMP gene.

Antibodies

Due to its structural similarity with the ARMP, the E5-1 protein may be used for the development of probes, peptides, or antibodies to various peptides within it which may recognize both the E5-1 and the ARMP gene and gene products,

respectively. As a protein homologue for the ARMP, the E5-1 protein may be used as a replacement for a defective ARMP gene product. It may also be used to elucidate functions of the ARMP gene in tissue culture and vice versa.

5 Screening for Alzheimer's Disease linked to Chromosome 1

Screening for Alzheimer's Disease linked to mutations of the *E5-1* gene may now be conveniently carried out.

General screening methods are described above in relation to the described mutations in the ARMP gene. These described methods can be readily applied and adapted to detection of the described chromosome 1 mutations, as will be readily understood by those skilled in the art.

In accordance with one embodiment of the invention, the Asn141Ile mutation is screened for by PCR amplification of the surrounding DNA fragment using the primers:

15 1041: 5'-cattcactgaggacacacc (end-labelled) and
 1042: 5'-tgtagagcaccaccaaga (unlabelled)

Any tissue with nucleated cell may be examined. The amplified products are separated by electrophoresis and an autoradiogram of the gel is prepared and examined for mutant bands.

20 In accordance with a further embodiment, the Met239Val mutation is screened for by PCR amplification of its surrounding DNA fragment using the primers:

 1034: 5'-gcatggtgtgcatccact and
 1035: 5'-ggaccactctgggaggtg

25 The amplified products are separated and an autoradiogram prepared as described above to detect mutant bands.

The same primer sets may be used to detect the mutations by means of other methods such as SSCP, chemical cleavage, DGGE, nucleotide sequencing, ligation chain reaction and allele specific oligonucleotides. As will be understood by those skilled in the art, other suitable primer pairs may be devised and used.

30 In inherited cases, as the primary event, and in non-inherited cases as a

secondary event due to the disease state, abnormal processing of E5-1, ARMP, APP or proteins reacting with E5-1, APP or ARMP, may occur. This can be detected as abnormal phosphorylation, glycosylation, glycation amidation or proteolytic cleavage products in body tissues or fluids, eg. CSF or blood.

5

Therapies

An important aspect of the biochemical studies using the genetic information of this invention is the development of therapies to circumvent or overcome the ARMP gene defect, and thus prevent, treat, control serious symptoms or cure the disease. In view of expression of the ARMP gene in a variety of tissues, one has to recognize that Alzheimer's Disease may not be restricted to the brain. Alzheimer's Disease manifests itself as a neurological disorder which in one of its forms is caused by a mutation in the ARMP gene, but such manifest may be caused by the mutations in other organ tissues, such as the liver, releasing factors which affect the brain activity and ultimately cause Alzheimer's Disease. Hence, in considering various therapies, it is understood that such therapies may be targeted at tissue other than the brain, such as heart, placenta, lung, liver, skeletal muscle, kidney and pancreas, where ARMP is also expressed.

The effect of these mutations in E5-1 and ARMP is a gain of a novel function which causes aberrant processing of (APP) Amyloid Precursor Protein into A β peptide, abnormal phosphorylation homeostasis, and abnormal apoptosis. Therapy to reverse this will be small molecules (drugs) recombinant proteins, etc. which block the aberrant function by altering the structure of the mutant protein, enhancing its metabolic clearance or inhibiting binding of ligands to the mutant protein, or inhibiting the channel function of the mutant protein. The same effect might be gained by inserting a second mutant protein by gene therapy similar to the correction of the "Deg 1(d)" and "Mec 4(d)" mutations in *C. elegans* by insertion of mutant transgenes. Alternately overexpression of wild type E5-1 protein or wild type ARMP or both may correct the defect. This could be the administration of drugs or proteins to induce the transcription and translation or inhibit the catabolism of the native E5-1

and ARMP proteins. It could also be accomplished by infusion of recombinant proteins or by gene therapy with vectors causing expression of the normal protein at a high level.

5 Rationale for Therapeutic, Diagnostic, and Investigational Applications of the ARMP
 Gene and Gene Products as They Relate to the Amyloid Precursor Protein

 The A β peptide derivatives of APP are neurotoxic (Selkoe et al, 1994). APP
is metabolized by passages through the Golgi network and then to secretory pathways
via clathrin-coated vesicles with subsequent passage to the plasma membrane where
10 the mature APP is cleaved by α -secretase to a soluble fraction (Protease Nexin II)
 plus a non-amyloidogenic C-terminal peptide (Selkoe et al. 1995, Gandy et al. 1993).
 Alternatively, mature APP can be directed to the endosome-lysosome pathway where
it undergoes beta and gamma secretase cleavage to produce the A β peptides. The
phosphorylation state of the cell determines the relative balance of α -secretase
15 (non-amyloidogenic) or A β pathways (amyloidogenic pathway) (Gandy et al. 1993).
 The phosphorylation state of the cell can be modified pharmacologically by phorbol
esters, muscarinic agonists and other agents, and appears to be mediated by cytosolic
factors (especially protein kinase C) acting upon an integral membrane protein in the
Golgi network, which we propose to be the ARMP, and members of the homologous
20 family (all of which carry several phosphorylation consensus sequences for protein
 kinase C). Mutations in the ARMP gene will cause alterations in the structure and
function of the ARMP gene product leading to defective interactions with regulatory
elements (eg. protein kinase C) or with APP, thereby promoting APP to be directed
to the amyloidogenic endosome-lysosome pathway. Environmental factors (viruses,
25 toxins, and aging etc) may also have similar effects on ARMP. To treat Alzheimer's
 disease, the phosphorylation state of ARMP can be altered by chemical and
biochemical agents (eg. drugs, peptides and other compounds) which alter the activity
of protein kinase C and other protein kinases, or which alter the activity of protein
phosphatases, or which modify the availability of ARMP to be postrationally
30 modified. The interactions between kinases and phosphatases with the ARMP gene

products (and the products of its homologues), and the interactions of the ARMP gene products with other proteins involved in the trafficking of APP within the Golgi network can be modulated to decrease trafficking of Golgi vesicles to the endosome-lysosome pathway thereby promoting A β peptide production. Such compounds will include: peptide analogues of APP, ARMP, and homologues of ARMP as well as other interacting proteins, lipids, sugars, and agents which promote differential glycosylation of ARMP and its homologues; agents which alter the biologic half-life of messenger RNA or protein of ARMP and homologues including antibodies and antisense oligonucleotides; and agents which act upon ARMP transcription.

The effect of these agents in cell lines and whole animals can be monitored by monitoring: transcription; translation; post-translational modification of ARMP (eg phosphorylation or glycosylation); and intracellular trafficking of ARMP and its homologues through various intracellular and extracellular compartments. Methods for these studies include Western and Northern blots; immunoprecipitation after metabolic labelling (pulse-chase) with radio-labelled methionine and ATP, and immunohistochemistry. The effect of these agents can also be monitored using studies which examine the relative binding affinities and relative amounts of ARMP gene products involved in interactions with protein kinase C and/or APP using either standard binding affinity assays or co-precipitation and Western blots using antibodies to protein kinase C, APP or ARMP and its homologues. The effect of these agents can also be monitored by assessing the production of A β peptides by ELISA before and after exposure to the putative therapeutic agent (Huang et al. 1993). The effect can also be monitored by assessing the viability of cell lines after exposure to aluminum salts and to A β peptides which are thought to be neurotoxic in Alzheimer's disease. Finally, the effect of these agents can be monitored by assessing the cognitive function of animals bearing: their normal genotype at APP or ARMP homologues; or bearing human APP transgenes (with or without mutations); or bearing human ARMP transgenes (with or without mutations); or a combination of all of these.

Rationale for Therapeutic, Diagnostic, and Investigational Applications of the ARMP Gene, the *E5-1* gene and their products

The ARMP gene product and the *E5-1* gene product have amino acid sequence homology to human ion channel proteins and receptors. For instance, the *E5-1* protein shows substantial homology to the human sodium channel α -subunit ($E=0.18$, $P=0.16$, identities = 22 - 27% over two regions of at least 35 amino acid residues) using the BLASTP paradigm of Altschul et al. 1990. Other diseases (such as malignant hyperthermia and hyperkalemic periodic paralysis in humans and the neurodegenerative of mechanosensory neurons in *C. elegans*) arise through mutations in ion channels or receptor proteins. Mutation of the ARMP gene or the *E5-1* gene could affect similar functions and lead to Alzheimer's disease and other psychiatric and neurological diseases. Based upon this, a test for Alzheimer's disease can be produced to detect an abnormal receptor or an abnormal ion channel function related to abnormalities that are acquired or inherited in the ARMP gene and its product, or in one of the homologous genes such as *E5-1* and their products. This test can be accomplished either *in vivo* or *in vitro* by measurements of ion channel fluxes and/or transmembrane voltage or current fluxes using patch clamp, voltage clamp and fluorescent dyes sensitive to intracellular calcium or transmembrane voltage. Defective ion channel or receptor function can also be assayed by measurements of activation of second messengers such as cyclic AMP, cGMP tyrosine kinases, phosphates, increases in intracellular Ca^{2+} levels, etc. Recombinantly made proteins may also be reconstructed in artificial membrane systems to study ion channel conductance. Therapies which affect Alzheimer's disease (due to acquired/inherited defects in the ARMP gene or *E5-1* gene; due to defects in other pathways leading to this disease such as mutations in APP; and due to environmental agents) can be tested by analysis of their ability to modify an abnormal ion channel or receptor function induced by mutation in the ARMP gene or in one of its homologues. Therapies could also be tested by their ability to modify the normal function of an ion channel or receptor capacity of the ARMP gene products and its homologues. Such assays can be performed on cultured cells expressing endogenous normal or mutant ARMP

genes/gene products or *E5-1* genes/gene products. Such studies can be performed in addition on cells transfected with vectors capable of expressing ARMP, parts of the ARMP gene and gene product, mutant ARMP, *E5-1* gene, parts of the *E5-1* gene and gene product, mutant *E5-1* gene or another homologue in normal or mutant form.

- 5 Therapies for Alzheimer's disease can be devised to modify an abnormal ion channel or receptor function of the ARMP gene or *E5-1* gene. Such therapies can be conventional drugs, peptides, sugars, or lipids, as well as antibodies or other ligands which affect the properties of the ARMP or *E5-1* gene product. Such therapies can also be performed by direct replacement of the ARMP gene and/or *E5-1* gene by gene
- 10 therapy. In the case of an ion channel, the gene therapy could be performed using either mini-genes (cDNA plus a promoter) or genomic constructs bearing genomic DNA sequences for parts or all of the ARMP gene. Mutant ARMP or homologous gene sequences might also be used to counter the effect of the inherited or acquired abnormalities of the ARMP gene as has recently been done for replacement of the
- 15 *mec 4* and *deg 1* in *C.elegans* (Huang and Chalfie, 1994). The therapy might also be directed at augmenting the receptor or ion channel function of the homologous genes such as the *E5-1* gene, in order that it may potentially take over the functions of the ARMP gene rendered defective by acquired or inherited defects. Therapy using antisense oligonucleotides to block the expression of the mutant ARMP gene
- 20 or the mutant *E5-1* gene, coordinated with gene replacement with normal ARMP or *E5-1* gene can also be applied using standard techniques of either gene therapy or protein replacement therapy.

Protein Therapy

- 25 Treatment of Alzheimer's Disease can be performed by replacing the mutant protein with normal protein, or by modulating the function of the mutant protein. Once the biological pathway of the ARMP protein has been completely understood, it may also be possible to modify the pathophysiologic pathway (eg. a signal transduction pathway) in which the protein participates in order to correct the
- 30 physiological defect.

To replace the mutant protein with normal protein, or with a protein bearing a deliberate counterbalancing mutation it is necessary to obtain large amounts of pure ARMP protein or E5-1 protein from cultured cell systems which can express the protein. Delivery of the protein to the affected brain areas or other tissues can then be accomplished using appropriate packaging or administering systems.

Gene Therapy

Gene therapy is another potential therapeutic approach in which normal copies of the ARMP gene are introduced into patients to successfully code for normal protein in several different affected cell types. The gene must be delivered to those cells in a form in which it can be taken up and code for sufficient protein to provide effective function. Alternatively, in some neurologic mutants it has been possible to prevent disease by introducing another copy of the homologous gene bearing a second mutation in that gene or to alter the mutation, or use another gene to block its effect.

Retroviral vectors can be used for somatic cell gene therapy especially because of their high efficiency of infection and stable integration and expression. The targeted cells however must be able to divide and the expression of the levels of normal protein should be high because the disease is a dominant one. The full length ARMP gene can be cloned into a retroviral vector and driven from its endogenous promoter or from the retroviral long terminal repeat or from a promoter specific for the target cell type of interest (such as neurons).

Other viral vectors which can be used include adeno-associated virus, vaccinia virus, bovine papilloma virus, or a herpesvirus such as Epstein-Barr virus.

Gene transfer could also be achieved using non-viral means requiring infection *in vitro*. This would include calcium phosphate, DEAE dextran, electroporation, and protoplast fusion. Liposomes may also be potentially beneficial for delivery of DNA into a cell. Although these methods are available, many of these are lower efficiency.

Antisense based strategies can be employed to explore ARMP gene function and as a basis for therapeutic drug design. The principle is based on the hypothesis

that sequence-specific suppression of gene expression can be achieved by intracellular hybridization between mRNA and a complementary antisense species. The formation of a hybrid RNA duplex may then interfere with the processing/transport/translation and/or stability of the target ARMP mRNA. Hybridization is required for the antisense effect to occur, however the efficiency of intracellular hybridization is low and therefore the consequences of such an event may not be very successful. Antisense strategies may use a variety of approaches including the use of antisense oligonucleotides, injection of antisense RNA and transfection of antisense RNA expression vectors. Antisense effects can be induced by control (sense) sequences, however, the extent of phenotypic changes are highly variable. Phenotypic effects induced by antisense effects are based on changes in criteria such as protein levels, protein activity measurement, and target mRNA levels. Multidrug resistance is a useful model to study molecular events associated with phenotypic changes due to antisense effects, since the multidrug resistance phenotype can be established by expression of a single gene *mdr1* (MDR gene) encoding for P-glycoprotein.

Transplantation of normal genes into the affected area of the patient can also be useful therapy for Alzheimer's Disease. In this procedure, a normal hARMP protein is transferred into a cultivatable cell type such as glial cells, either exogenously or endogenously to the patient. These cells are then injected serotologically into the disease affected tissue(s). This is a known treatment for Parkinson's disease.

Immunotherapy is also possible for Alzheimer's Disease. Antibodies can be raised to a mutant ARMP protein (or portion thereof) and then be administered to bind or block the mutant protein and its deleterious effects. Simultaneously, expression of the normal protein product could be encouraged. Administration could be in the form of a one time immunogenic preparation or vaccine immunization. An immunogenic composition may be prepared as injectables, as liquid solutions or emulsions. The ARMP protein may be mixed with pharmaceutically acceptable excipients compatible with the protein. Such excipients may include water, saline, dextrose, glycerol, ethanol and combinations thereof. The immunogenic composition

and vaccine may further contain auxiliary substances such as emulsifying agents or adjuvants to enhance effectiveness. Immunogenic compositions and vaccines may be administered parenterally by injection subcutaneously or intramuscularly.

5 The immunogenic preparations and vaccines are administered in such amount as will be therapeutically effective, protective and immunogenic. Dosage depends on the route of administration and will vary according to the size of the host.

Similar gene therapy techniques may be employed with respect to the *E5-1* gene.

10 The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples. These examples are described solely for purposes of illustration and are not intended to limit the scope of the invention. Changes in the form and substitution of equivalents are contemplated as circumstances may suggest or render expedient. Although specific terms have been employed herein, such terms are intended in a
15 descriptive sense and not for purposes of limitations.

Example 1. Development of the Genetic, physical "contig" and transcriptional map of the minimal co-segregating region

The CEPH MegaYAC and the RPCI PAC human total genomic DNA libraries were searched for clones containing genomic DNA fragments from the AD3 region of chromosome 14q24.3 using oligonucleotide probes for each of the ## SSR marker loci used in the genetic linkage studies as well as ## additional markers depicted in Figure 1a (Albertsen et al., 1990; Chumakov et al., 1992; Ioannu et al., 1994). The genetic map distances between each marker are depicted above the contig, and are derived from published data (NIH/CEPH Collaborative Mapping Group, 1992; Wang, 1992; Weissenbach, J et al., 1992; Gyapay, G et al., 1994). Clones recovered for each of the initial marker loci were arranged into an ordered series of partially overlapping clones ("contig") using four independent methods. First, sequences representing the ends of the YAC insert were isolated by inverse PCR (Riley et al., 1990), and hybridized to Southern blot panels containing restriction digests of DNA from all of the YAC clones recovered for all of the initial loci in order to identify other YAC clones bearing overlapping sequences. Second, inter-Alu PCR was performed on each YAC, and the resultant band patterns were compared across the pool of recovered YAC clones in order to identify other clones bearing overlapping sequences (Bellamne-Chartelot et al., 1992; Chumakov et al., 1992). Third, to improve the specificity of the Alu-PCR fingerprinting, we restricted the YAC DNA with HaeIII or RsaI, amplified the restriction products with both Alu and L1H consensus primers, and resolved the products by polyacrylamide gel electrophoresis. Finally, as additional STSs were generated during the search for transcribed sequences, these STSs were also used to identify overlaps. The resultant contig was complete except for a single discontinuity between YAC932C7 bearing *D14S53* and YAC746B4 containing *D14S61*. The physical map order of the STSs within the contig was largely in accordance with the genetic linkage map for this region (NIH/CEPH Collaborative Mapping Group, 1992; Wang, Z, Weber, J.L., 1992; Weissenbach, J et al., 1992; Gyapay, G et al., 1994). However, as with the genetic maps, we were unable to unambiguously resolve the relative order of the loci within the

D14S43/D14S71 cluster and the *D14S76/D14S273* cluster. PAC1 clones suggest that *D14S277* is telomeric to *D14S268*, whereas genetic maps have suggested the reverse order. Furthermore, a few STS probes failed to detect hybridization patterns in at least one YAC clone which, on the basis of the most parsimonious consensus physical map and from the genetic map, would have been predicted to contain that STS. For instance, the *D14S268* (AFM265) and RSCAT7 STSs are absent from YAC788H12 (Figure 3). Because these results were reproducible, and occurred with several different STS markers, these results most likely reflect the presence of small interstitial deletions within one of the YAC clones.

Example 2. Cumulative two-point lod scores for chromosome 14q24.3 markers.

Genotypes at each polymorphic microsatellite marker locus were determined by PCR from 100ng of genomic DNA of all available affected and unaffected pedigree members as previously described (St George-Hyslop, P et al., 1992) using primer sequences specific for each microsatellite locus (Weissenbach, J et al., 1992; Gyapay, G et al., 1994). The normal population frequency of each allele was determined using spouses and other neurologically normal subjects from the same ethnic groups, but did not differ significantly from those established for mixed Caucasian populations (Weissenbach, J et al., 1992; Gyapay, G et al., 1994). The maximum likelihood calculations assumed an age of onset correction, marker allele frequencies derived from published series of mixed Caucasian subjects, and an estimated allele frequency for the AD3 mutation of 1:1000 as previously described (St George-Hyslop, P et al., 1992). The analyses were repeated using equal marker allele frequencies, and using phenotype information only from affected pedigree members as previously described to ensure that inaccuracies in the estimated parameters used in the maximum likelihood calculations did not misdirect the analyses (St George-Hyslop, P et al., 1992). These supplemental analyses did not significantly alter either the evidence supporting linkage, or the discovery of recombination events.

Example 3. Haplotypes between flanking markers segregated with AD3 in FAD pedigrees

Extended haplotypes between the centromeric and telomeric flanking markers on the parental copy of chromosome 14 segregating with AD3 in fourteen early onset FAD pedigrees (pedigrees NIH2, MGH1, Tor1.1, FAD4, FAD1, MEX1, and FAD2 show pedigree specific lod scores $\geq +3.00$ with at least one marker between D14S258 and D14S53). Identical partial haplotypes (boxed) are observed in two regions of the disease bearing chromosome segregating in several pedigrees of similar ethnic origin. In region A, shared alleles are seen at D14S268 ("B": allele size = 126 bp, allele frequency in normal Caucasians = 0.04; "C": size = 124 bp, frequency = 0.38); D14S277 ("B": size = 156 bp, frequency = 0.19; "C": size = 154 bp, frequency = 0.33); and RSCAT6 ("D": size = 111bp, frequency 0.25; "E": size = 109bp, frequency = 0.20; "F": size = 107 bp, frequency = 0.47). In region B, alleles of identical size are observed at D14S43 ("A": size = 193bp, frequency = 0.01; "D": size = 187 bp, frequency = 0.12; "E": size = 185 bp, frequency = 0.26; "I": size = 160 bp, frequency = 0.38); D14S273 ("3": size = 193 bp, frequency = 0.38; "4" size = 191 bp, frequency = 0.16; "5": size = 189 bp, frequency = 0.34; "6": size = 187 bp, frequency = 0.02) and D14S76 ("1": size = bp, frequency = 0.01; "5": size = bp, frequency = 0.38; "6": size = bp, frequency = 0.07; "9": size = bp, frequency = 0.38). The ethnic origins of each pedigree are abbreviated as: Ashk = Ashkenazi Jewish; Ital = Southern Italian; Angl = Anglo-Saxon-Celt; FrCan = French Canadian; Jpn = Japanese; Mex = Mexican Caucasian; Ger = German; Am = American Caucasian. The type of mutation detected is depicted by the amino acid substitution and putative codon number or by ND where no mutation has been detected because a comprehensive survey has not been undertaken due to the absence of a source of mRNA for RT-PCR studies.

Example 4. Recovery of transcribed sequences from the AD3 interval.

Putative transcribed sequences encoded in the AD3 interval were recovered using either a direct hybridization method in which short cDNA fragments generated

from human brain mRNA were hybridized to immobilized cloned genomic DNA fragments (Rommens, JM et al., 1993). The resultant short putatively transcribed sequences were used as probes to recover longer transcripts from human brain cDNA libraries (Stratagene, La Jolla). The physical location of the original short clone and of the subsequently acquired longer cDNA clones were established by analysis of the hybridization pattern generated by hybridizing the probe to Southern blots containing a panel of EcoRI digested total DNA samples isolated from individual YAC clones within the contig. The nucleotide sequence of each of the longer cDNA clones was determined by automated cycle sequencing (Applied Biosystems Inc., CA), and compared to other sequences in nucleotide and protein databases using the blast algorithm (Altschul, SF et al., 1990). Accession numbers for the transcribed sequences in this report are: L40391, L40392, L40393, L40394, L40395, L40396, L40397, L40398, L40399, L40400, L40401, L40402, and L40403.

Example 5. Locating mutations in the ARMP gene using restriction enzymes.

The presence of Ala 246 Glu mutation which creates a DdeI restriction site was assayed in genomic DNA by PCR using the end labelled primer 849 (5'-atctccggcaggcatatct-3') SEQ ID No:126 and the unlabelled primer 892 (5'-tgaaatcacagccaagatgag-3') SEQ ID No:127 to amplify an 84bp genomic exon fragment using 100ng of genomic DNA template, 2mM MgCl₂, 10 pMoles of each primer, 0.5U Taq polymerase, 250 uM dNTPs for 30 cycles of 95°C X 20 seconds, 60°C X 20 seconds, 72°C X 5 seconds. The products were incubated with an excess of DdeI for 2 hours according to the manufacturers protocol, and the resulting restriction fragments were resolved on a 6% nondenaturing polyacrylamide gel and visualized by autoradiography. The presence of the mutation was inferred from the cleavage of the 84bp fragment to due to the presence of a DdeI restriction site. All affected members of the FAD1 pedigree (filled symbols) and several at-risk members ("R") carried the DdeI site. None of the obligate escapees (those individuals who do not get the disease, age > 70years), and none of the normal controls carried the DdeI mutation.

Example 6. Locating mutation in the ARMP gene using allele specific oligonucleotides.

The presence of the Cys 410 Tyr mutation was assayed using allele specific oligonucleotides. 100ng of genomic DNA was amplified with the exonic sequence primer 885 (5'-tggagactggaacacaac-3') SEQ ID No:128 and the opposing intronic sequence primer 893 (5'-gtgtggccagggtagagaact-3') SEQ ID No:129 using the above reaction conditions except 2.5 mM MgCl₂, and cycle conditions of 94°C X 20 seconds, 58°C X 20 seconds, and 72°C for 10 seconds). The resultant 216bp genomic fragment was denatured by 10-fold dilution in 0.4M NaOH, 25 mM EDTA, and was vacuum slot-botted to duplicate nylon membranes. The end-labelled "wild type" primer 890 (5'-ccatagcctgtttcgtagc-3') Seq ID No:130 and the end-labelled "mutant" primer 891 (5'-ccatagcctAtttcgtagc-3') SEQ ID No:131 were hybridized to separate copies of the slot-blot filters in 5 X SSC, 5 X Denhardt's, 0.5% SDS for 1 hour at 48°C, and then washed successively in 2 X SSC at 23°C and 2 X SSC, 0.1% SDS at 50°C and then exposed to X-ray film. All testable affected members as well as some at-risk members of the AD3 (shown) and NIH2 pedigrees (not shown) possessed the Cys 410 Tyr mutation. Attempts to detect the Cys 410 Tyr mutation by SSCP revealed that a common intronic sequence polymorphism migrated with the same SSCP pattern.

Example 7. Northern hybridization demonstrating the expression of ARMP protein mRNA in a variety of tissues.

Total cytoplasmic RNA was isolated from various tissue samples (including heart, brain and different regions of, placenta, lung, liver, skeletal muscle, kidney and pancreas) obtained from surgical pathology using standard procedures such as CsCl purification. The RNA was then electrophoresed on a formaldehyde gel to permit size fractionation. The nitrocellulose membrane was prepared and the RNA was then transferred onto the membrane. ³²P-labelled cDNA probes were prepared and added to the membrane in order for hybridization between the probe the RNA to occur. After washing, the membrane was wrapped in plastic film and placed into

imaging cassettes containing X-ray film. The autoradiographs were then allowed to develop for one to several days. The positions of the 28S and 18S rRNA bands are indicated. Sizing was established by comparison to standard RNA markers. Analysis of the autoradiographs revealed a prominent band at 3.0 kb in size. These northern blots demonstrated the ARMP gene is expressed in all of the tissues examined.

Example 8: Eukaryotic and Prokaryotic Expression Vector Systems

Eukaryotic and prokaryotic expression systems have been generated using two different classes of ARMP nucleotide cDNA sequence inserts. In the first class, termed full-length constructs, the entire ARMP cDNA sequence was inserted into the expression plasmid in the correct orientation, and included both the natural 5' UTR and 3' UTR sequences as well as the entire open reading frame. The open reading frames bear a nucleotide sequence cassette which allows either the wild type open reading frame to be included in the expression system or alternatively, single or a combination of double mutations can be inserted into the open reading frame. This was accomplished by removing a restriction fragment from the wild type open reading frame using the enzymes NarI and PflmI and replacing it with a similar fragment generated by reverse transcriptase PCR and which bears the nucleotide sequence encoding either the Met146Leu mutation or the Hys163Arg mutation. A second restriction fragment was removed from the wild type normal nucleotide sequence for the open reading frame by cleavage with the enzymes PflmI and NcoI and replaced with restriction fragments bearing with the nucleotide sequence encoding the Ala246Glu mutation, or the Ala260Val mutation or the Ala285Val mutation or the Leu286Val mutation, or the Leu392Val mutation, or the Cys410Tyr mutation. Finally, a third variant bearing combinations of either the Met146Leu or His163Arg mutations in tandem with the remaining mutations by linking the NarI-PflmI fragment bearing these mutations and the PflmI-NcoI fragments bearing the remaining mutations.

A second variant of cDNA inserts bearing wild type or mutant cDNA sequences was constructed by removing from the full-length cDNA the 5' UTR and

part of the 3' UTR sequences. The 5' UTR sequence was replaced with a synthetic oligonucleotide containing a KpnI restriction site and a Kozak initiation site (oligonucleotide 969: ggtaccgccaccatgacagagggtacctgcac) SEQ ID No:138. The 3' UTR was replaced with an oligonucleotide corresponding to position 2566 of the cDNA and bears an artificial EcoRI site (oligonucleotide 970:gaattcactggctgtagaaaaagac) SEQ ID No:139. Mutant variants of this construct were then made by inserting the same mutant sequences described above at the NarI-PflmI fragment, and at the PstmI-NcoI sites described above.

For eukaryotic expressions, these various cDNA constructs bearing wild type and mutant sequences were cloned into the expression vector pZeoSV (invitrogen). For prokaryotic expression, two constructs were made using the glutathione S-transferase fusion vector pGEX-kg. The inserts which have been attached to the GST fusion nucleotide sequence are the same nucleotide sequence described above generated with the oligonucleotide primers 969, Sequence ID No:138 and 970, Sequence ID No:139, bearing either the normal open reading frame nucleotide sequence or bearing a combination of single and double mutations as described above. This construct allows expression of the full-length protein in mutant and wild type variants in prokaryotic cell systems as a GST fusion protein which will allow purification of the full-length protein followed by removal of the GST fusion product by thrombin digestion. The second prokaryotic cDNA construct was generated to create a fusion protein with the same vector, and allows the production of the amino acid sequence corresponding to the hydrophilic acidic loop domain between TM6 and TM7 of the full-length protein, as either a wild type nucleotide sequence (thus a wild type amino acid sequence for fusion proteins) or as a mutant sequence bearing either the Ala285Val mutation, or the Leu286Val mutation, or the Leu392Val mutation. This was accomplished by recovering wild type or mutant sequence from appropriate sources of RNA using the oligonucleotide primers 989:ggatccgggtccacttcgtatgctg SEQ ID No:140, and 990:tttttgaattcttaggctatggtgtgttcca SEQ ID No:141. This allows cloning of the appropriate mutant or wild type nucleotide sequence corresponding to the hydrophilic acid loop domain at the BamHI and the EcoRI sites within the pGEX-

KG vector.

These prokaryotic expression systems allow the holo-protein or various important functional domains of the protein to be recovered as fusion proteins and then used for binding studies, structural studies, functional studies, and for the generation of appropriate antibodies.

Example 9: Identification of Three New Mutations in the ARMP Gene

Three novel mutations have been identified in subjects affected with early onset Alzheimer's disease. All of these mutations co-segregate with the disease, and are absent from at least 200 normal chromosomes. The three mutations are as follows: a substitution of C by T at position 1027 which results in the substitution of alanine 260 for valine; substitution of C by T at position 1102, which results in the substitution of alanine at 285 by valine; and substitution of C by G at position 1422 which results in the substitution of leucine 392 by valine. Significantly, all of these mutations occur within the acidic hydrophilic loop between putative TM6 and TM7. Two of the mutations (A260V; A285V) and the L286V mutation are also located in the alternative spliced domain.

The three new mutations, like the other mutations, can be assayed by a variety of strategies (direct nucleotide sequencing, Allele specific oligos, ligation polymerase chain reaction, SSCP, RFLPs) using RT-PCR products representing the mature mRNA/cDNA sequence or genomic DNA. We have chosen allele specific oligos. For the A260V and the A285V mutations, genomic DNA carrying the exon can be amplified using the same PCR primers and methods as for the L286V mutation. PCR products were then denatured and slot blotted to duplicate nylon membranes using the slot blot protocol described for the C410T mutation.

The Ala260Val mutation was scored on these blots by using hybridization with end-labeled allele-specific oligonucleotides corresponding to the wild type sequence (994:gattagtgggtgttttg) SEQ ID No:142 or the mutant sequence (995:gattagtggctgttttg) SEQ ID No:143 by hybridization at 48°C followed by a wash at 52°C in 3X SSC buffer containing 0.1% SDS. The Ala285Val mutation was

scored on these slot blots as described above but using instead the allele-specific oligonucleotides for the wild type sequence (1003:ttttccagctctcattta) SEQ ID No:144 or the mutant primer (1004:ttttccagttctcattta) SEQ ID No:145 at 48°C followed by washing at 52°C as above except that the wash solution was 2X SSC.

5 The Leu392Val mutation was scored by amplification of the exon from genomic DNA using primers 996(aaacttgattgggagat) SEQ ID No:147 and 893 (gtgtggccagggtagagaact) SEQ ID No:129 using standard PCR buffer conditions excepting that the magnesium concentration was 2mM and cycle conditions were 94°C time 10 seconds, 56°C times 20 seconds, and 72°C for 10 seconds. The result
10 200 based pair genomic fragment was denatured as described for the Cys410Tyr mutation and slot-blotted in duplicate to nylon membranes. The presence or absence of the mutation was then scored by differential hybridization to either a wild type end-labelled oligonucleotide (999:tacagtgtctgttggtgta) SEQ ID No:146 or with an end-labeled mutant primer (100:tacagtgtgtgttggtgta) SEQ ID No:148 by hybridization at
15 45°C and then successive washing in 2X SSC at 23°C and then at 68°C.

Example 10: Polyclonal Antibody Production

 Peptide antigens were synthesized by solid-phase techniques and purified by reverse phase high pressure liquid chromatography. Peptides were covalently linked
20 to keyhole limpet hematoxylin (KLH) via disulfide linkages that were made possible by the addition of a cystein residue at the peptide C-terminus. This additional residue does not appear normally in the protein sequence and was included only to facilitate linkage to the KLH molecule. A total of three rabbits were immunized with peptide-KLH complexes for each peptide antigen and were then subsequently given booster
25 injections at seven day intervals. Antisera were collected for each peptide and pooled and IgG precipitated with ammonium sulfate. Antibodies were then affinity purified with Sulfo-link agarose (Pierce) coupled with the appropriate peptide. This final purification is required to remove non-specific interactions of other antibodies present in either the pre- or post- immune serum.

30 The specific sequences to which we have raised antibodies are;

Polyclonal antibody 1: NDNRERQEHNDRRSL (C)- residues 30-45
 Polyclonal antibody 2: KDGQLIYTPFTEDTE (C)- residues 109-120
 Polyclonal antibody 3: EAQRRVSKNSKYNAE (C)-residues 304-319
 Polyclonal antibody 4: SHLGPHRSTPESRAA (C)-residues 346-360

5 The non-native cysteine residue is indicated at the C-terminal by (C). These sequences are contained within various predicted domains of the protein. For example, antibodies 1, 3, and 4 are located in potentially functional domains that are exposed to the aqueous media and may be involved in binding to other proteins critical for the development of the disease phenotype. Antibody 2 corresponds to a
 10 short linking region situated between the predicted first and second transmembrane helices.

Example 11: Identification of two mutations in E5-1 gene

RT-PCR products corresponding to the *E5-1* ORF were generated from RNA
 15 of lymphoblasts or frozen post-mortem brain tissue using oligonucleotide primer pairs 1021:5'-cagaggatggagagaatac and 1018:5'-ggctcccaaaactgtcat (product = 888 bp); and 1071:5'-gccctagtgttcataagta and 1022: 5'-aaagcgggagccaaagtc (product = 826 bp) by PCR using 250 μ Mol dNTPs, 2.5 mM MgCl₂, 10 pMol oligonucleotides in 10 μ l cycled for 40 cycles of 94°C X 20 seconds, 58°C X 20 seconds, 72°C X 45
 20 seconds. The PCR products were sequenced by automated cycle sequencing (ABI, Foster City, A) and the fluorescent chromatograms were scanned for heterozygous nucleotide substitutions by direct inspection and by the Factura (ver 1.2.0) and Sequence Navigator (ver 1.0.1b15) software packages (data not shown).

Asn141Ile: The A→T substitution at nucleotide 787 creates a BclI restriction
 25 site. The exon bearing this mutation was amplified from 100 ng of genomic DNA using 10pMol of oligonucleotides 1041: 5'-cattcactgaggacacacc (end-labelled) and 1042: 5'-tgtagagcaccaccaaga (unlabelled), and PCR reaction conditions similar to those described below for the Met239Val. 2 μ l of the PCR product was restricted with BclI (NEBL, Beverly, MA) in 10 μ l reaction volume according to the
 30 manufacturers' protocol, and the products were resolved by non-denaturing

polyacrylamide gel electrophoresis. In subjects with wild type sequences, the 114 bp PCR product is cleaved into 68 bp and 46 bp fragments. Mutant sequences cause the product to be cleaved into 53 bp, 46 bp and 15 bp.

5 Met239Val: The A→G substitution at nucleotide 1080 deletes a NlaIII restriction site, allowing the presence of the Met 239Val mutation to be detected by amplification from 100 ng of genomic DNA using PCR (10 pMol oligonucleotides 1034: 5'-gcatggtgtgcatccact, 1035: 5'-ggaccactctgggaggta; 0.5 U Taq polymerase, 250 μM dNTPS, 1 μCi alpha ³²P-dCTP, 1.5 mM MgCl₂, 10 μl volume; 30 cycles of 94°C X 30 seconds, 58°C X 20 seconds, 72°C X 20 seconds) to generate a 110 bp
10 product. 2 μl of the PCR reaction were diluted to 10 μl and restricted with 3 U of NlaIII (NEBL, Beverly, MA) for 3 hours. The restriction products were resolved by non-denaturing polyacrylamide gel electrophoresis and visualized by autoradiography. Normal subjects show cleavage products of 55, 35, 15 and 6bp, whereas the mutant sequence gives fragments of 55, 50 and 6 bp.

15 Although preferred embodiments of the invention have been described herein in detail, it will be understood by those skilled in the art that variations may be made thereto without departing from the spirit of the invention or the scope of the appended claims.

LOCUS	RECOMBINATION FRACTION (%)						
	0.00	0.05	0.10	0.15	0.20	0.25	0.40
D14S63	6	1.54	3.50	4.38	4.13	2.71	1.08
D14S258	6	21.60	19.64	17.19	14.50	8.97	3.31
D14S77	6	15.18	15.53	14.35	12.50	7.82	2.92
D14S71	6	15.63	14.14	12.19	10.10	5.98	2.39
D14S43	6	19.36	17.31	15.27	12.94	7.30	3.11
D14S273	6	12.10	11.52	10.12	8.48	5.04	1.91
D14S61	6	26.90	24.92	22.14	18.38	12.05	5.07
D14S33	6	11.32	11.41	10.39	8.99	5.73	2.51
D14S48	6	0.50	1.03	1.14	1.04	0.60	0.13

TABLE 1

PEDIGREE ID

LOCUS	NIH2	FAD1	TUR1.1	FAD4	RE	FAD1	MG12	BOW	COOK	683	Tor42	QUE	MEX1	FAD2
D14863	1	4	7	4		5							9	2
D148258	6	6	8	7	4	5	5	6	6		7	6	7	6
D148268	C	C	B	B	C	C	C	C	C	C	C	C	C	C
D148277	C	C	C	C	C	C	C	C	C	C	C	C	C	C
D148736	D	D	E	E	F	E	E	D/F	E	E	E	E	F	D
D14877	Y	Y	K	S		P	P	K	H		C	U	F	A
D14871	7	7	1	2	7	7		6	7		3	7	2	6
D14843	A	A	1	1	1	E	D	1	1		C	1	D	C
D148273	6	6	3	5	5	4	4	4	6		6	6	5	5
D14876	5	5	5	5	5	6	9	9			9	1	5	5
D14861	E	E	G	F		1					D		L	F
D14853	F	F	C	F	F	J	C	F	E		J	D	F	F
ETHNIC ORIGIN	Ashk	Ashk	Ital	Ital	Ital	Angl	Angl	Angl	Angl	Amer	FrCan	FrCan	Mex	Ger
MUTATION	C410Y	C410Y	M146L	M146L	ND	A246E	ND	ND	ND	H163R	H163R	ND	ND	L236V

TABLE 2

HUMAN
MUSE

No.	Target File	Key	Target	Overlap	Match	Percentage	Similarity
1	camp.com/long (Frame 1)	1	1	467	465	99.57%	
1	10	20	30	40	50	60	70
N-	N- MTELPAFLSYTCQACMSHMLARTVRSQCHRRRQSHDRRLGTFPLNKRPCQNHQVVEQDEED						
N-	N- MTELPAFLSYTCQACMSHMLARTVRSQCHRRRQSHDRRLGTFPLNKRPCQNHQVVEQDEED						
1	10	20	30	40	50	60	70
71	80	90	100	110	120	130	140
SEMLXGCAKIVDLFVPTLQVWVATKGVSTYTRDGLTTFPTTETVQGRALSTLRAALMI							
SEMLXGCAKIVDLFVPTLQVWVATKGVSTYTRDGLTTFPTTETVQGRALSTLRAALMI							
71	80	90	100	110	120	130	140
141	150	160	170	180	190	200	210
SVTVVMTLLVVLVYKRCXVTHAWLINSILLTFFPTLQVWVATKGVSTYTRDGLTTFPTTETVQGRALSTLRAALMI							
SVTVVMTLLVVLVYKRCXVTHAWLINSILLTFFPTLQVWVATKGVSTYTRDGLTTFPTTETVQGRALSTLRAALMI							
141	150	160	170	180	190	200	210
211	220	230	240	250	260	270	280
IEMWKPRLQCAVLEISALMALVFTKLPENTWLLAVISVVELVAVLCYKSTPLMLVETACERNE							
IEMWKPRLQCAVLEISALMALVFTKLPENTWLLAVISVVELVAVLCYKSTPLMLVETACERNE							
211	220	230	240	250	260	270	280
281	290	300	310	320	330	340	350
LTFPALINESTVWLVNMAEGDPEACRRVSEHETNAESTERESCDTVAENDCGFSSEVACACSHLGP							
LTFPALINESTVWLVNMAEGDPEACRRVSEHETNAESTERESCDTVAENDCGFSSEVACACSHLGP							
281	290	300	310	320	330	340	350
351	360	370	380	390	400	410	420
HRSTPESRAAVQELGSEITSEDPEERGVVLGACDFIFSVLVCKASNTSGENMTTRCKVALLIGLCL							
HRSTPESRAAVQELGSEITSEDPEERGVVLGACDFIFSVLVCKASNTSGENMTTRCKVALLIGLCL							
351	360	370	380	390	400	410	420
421	430	440	450	460	470	480	490
TILLALTYKCALPAPISITTFGLVTFYADYLVQFFACCLAFROFTI -C							
TILLALTYKCALPAPISITTFGLVTFYADYLVQFFACCLAFROFTI -C							
421	430	440	450	460	470	480	490

Table 3.

HUMAN ARMP FUNCTIONAL DOMAINS

Domains (Amino Acid Residue)				Functional Characteristic
82	-	100	AA	Hydrophobic
132	-	154	AA	Hydrophobic
164	-	183	AA	Hydrophobic
195	-	213	AA	Hydrophobic
221	-	238	AA	Hydrophobic
244	-	256	AA	Hydrophobic
281	-	299	AA	Hydrophobic
404	-	428	AA	Hydrophobic
431	-	449	AA	Hydrophobic
115	-	119	AA (YTPF)	Phosphorylation Site
353	-	356	AA (STPC)	Phosphorylation Site
300	-	385	AA	Acid Rich Domain
				Possible Metal Binding Domain

ANTIGENIC SITES INCLUDING AMINO ACID RESIDUES

27	-	44
46	-	48
50	-	60
66	-	67
107	-	111
120	-	121
125	-	126
155	-	160
185	-	189
214	-	223
220	-	230
240	-	245
267	-	269
273	-	282
300	-	370
400	-	420

TABLE 4

M145LPU	BspHI (destroy)	210 (170-S182 ID) TCACAGAAAGATA CCGAGACT	211 (170-S182) R CCCAACCATAAAG AGAACA	
MIS 163 Arg	Nla III (destroy)	227 (Intronle) TCTGTACTTTT AAGGTTTGTG	228 ACTTCAGAGTAATT CATCANCA	
Ala 246	Dlc I (create)	342 * GACTCCAGCAGG CATATCT	322 TGAAATCACAGCC AAGA'TGAG	
Leu 286 Val.	Pvu II (create)	232 GATGAGACAAGT NCCNTGAA 243 TTAGTGGCTGTT TNGTGTCC	251 CACCCATTACAAAG TTTAGC	
Cys 410 Tys	Allele specific iligo	383 GTGTGGCCAGGG TAGAGAACT	385 TGGAGACTGGAAC ACAAC	CCATAGCCCTGTTTCGTAGC 390=WT CCATAGCCCTATTTCGTAGC 391=MUT

TABLE 5

TABLE 6: POSITION OF EXONS AND INTRON-EXON BOUNDARIES OF THE ARMP GENE

cDNA/mRNA SEQUENCE		CORRESPONDING GENOMIC SEQUENCE	
ARMP (917 ver)	Transcript ID CC44 ver	Genomic sequences file ID & position of exon	Comments
1-113bp	N/A	917-936.gen @ 731-834bp	Alternate 5'UTR
N/A	1-422bp	917-936.gen @ 1067-1475bp	Alternate 5'UTR
114-195bp	423-500bp	932-943.gen @ 589-671bp	
196-335bp	501-632bp	932-943.gen @ 759-899bp	12bp Variably spliced
337-386bp	633-883bp	901-912.gen @ 787-1037bp	
587-730bp	884-1026bp	910-915.gen @ 1134-1278bp	M146L mutation
731-795bp	1027-1092bp	925-913.gen @ 413-478bp	H163R mutation
796-1017bp	1093-1314bp	849-892.gen @ 336-558bp	A246E mutation
1018-1116bp	1315-1413bp	951-952.gen @ 312-412bp	L286V mutation, variable spl
1117-1204bp	1414-1501bp	983-1011.gen @ 61-149bp	
1205-1377bp	1502-1674bp	874-984.gen @ 452-625bp	
1378-1497bp	1674-1794bp	885-1012.gen @ 431-550bp	C410Y mutation
1498-2760bp	1795-3060bp	930-919.gen @ ~10bp-end	last AA, STOP, 3'UTR

TABLE 7 : MUTATIONS AND POLYMORPHISMS IN THE ARMP GENE

Nucleotide # in ARMP.LPD	Amino acid # in ARMP.PRO	Comment
A->C ₃₄₄	Met146Leu	Pathogenic, Unique to AD affected.
A->G ₇₂₈	His163Arg	"
C->A ₇₄₃	Ala246Glu	"
C->T ₁₀₂₇	Ala260Val	"
C->T ₁₁₀₂	Ala235Val	"
C->G ₁₁₀₄	Leu236Val	"
C->G ₁₄₂₂	Leu392Val	"
G->A ₁₄₇₇	Cys410Tyr	"
G->T ₂₆₃	Phe205Leu	Polymorphism in normals
C->A ₁₇₀₀	non-coding	3'UTR polymorphism
G->A ₂₆₀₁	non-coding	"
delC ₂₅₁₃	non-coding	"

TABLE 8

[illegible]

SEQUENCE ID NO: 1

10	20	30	40	50	60
TGGGACAGGC	AGCTCGGGG	TCCCGCGTT	CACATCGAA	ACAAAACAGC	GGCTGGTCTG
70	80	90	100	110	120
GAAGGAACCT	GAGCTACGAG	CCCGCGGCG	AGCGGGGCG	CGGGGTAACG	GTATACCTAA
130	140	150	160	170	180
TCTGGGAGCC	TCCAAGTGAC	AACACCGTT	GGGTCTCTA	GACAGCTTGG	CCTGGAGGAG
190	200	210	220	230	240
AACACATGAA	AGAAAGAACC	TCAAGAGGCT	TTGTTTTCTG	TGAAACAGTA	TTTCTATACA
250	260	270	280	290	300
GTTGCTCCAA	TGACAGAGTT	ACCTGCACCG	TTGTCTACT	TCCAGAATGC	ACAGATGTCT
310	320	330	340	350	360
GAGGACAACC	ACCTGAGCAA	TACTGTACGT	AGCCAGGATG	ACAATAGAGA	ACGGCAGGAG
370	380	390	400	410	420
CACACGACA	GACCGAGCCT	TGGCCACCTT	GAGCCATTAT	CTAATGGACG	ACCCGAGGGT
430	440	450	460	470	480
AACTCCCGGC	AGGTGGTGGG	GCAAGATGAG	GAAGAAGATG	AGGAGGTCAC	ATTGAATAT
490	500	510	520	530	540
GGCCCGAACC	ATGTGATCAT	GGTCTTTGTC	CCTGTGACTC	TCTGCATGGT	GGTGGTCTGT
550	560	570	580	590	600
GCTACCATTA	AGTCAGTCAG	CTTTTATACC	CGGAAGGATG	GGCAGCTAAT	CTATACCCCA
610	620	630	640	650	660
TTACACAGAAG	ATACCGAGAC	TGTGGGCGAG	AGAGCCCTGC	ACTCAATTCT	GAATGCTGCC
670	680	690	700	710	720
ATCATGATCA	GTCTCATTTG	TGTATGACT	ATCCTCCTGG	TGGTTCTGTA	TAAATACGG
730	740	750	760	770	780
TGCTATAAGG	TCATCCATGC	CTGGCTTATT	ATATCATCTC	TATTGTGCT	GTCTTTTTT
790	800	810	820	830	840
TCATTCATTT	ACTTGGGGGA	AGTGTTTAAA	ACCTATAACG	TTGCTGTGGA	CTACATTACT
850	860	870	880	890	900
GTTGCACCTC	TGATCTGGAA	TTTGGGTGTC	GTGGGAATGA	TTTCCATTCA	CTGAAAAGGT
910	920	930	940	950	960
CCACTTCGAC	TCCAGCAGGC	ATATCTCATT	ATGATTAGTG	CCCTCATGGC	CCTGGTCTTT
970	980	990	1000	1010	1020
ATCAAGTACC	TCCCTGAATG	GACTGCGTGG	CTCATCTTGG	CTGTGATTTC	AGTATATGAT
1030	1040	1050	1060	1070	1080
TTAGTGGCTG	TTTTGTGTCC	GAAAGGTCCA	CTTCTATGCG	TGGTTGAAAC	AGCTCAGGAG

1090	1100	1110	1120	1130	1140
AGAAATGAAA	CGCTTTTTC	AGCTTTCATT	TACTCTTCAA	CAATCTCTTT	GTTCCTTCAAT
1150	1160	1170	1180	1190	1200
ATGCCAGAAG	GAGACCCGGA	AGCTCAAAGG	ACAGTATCCA	AAAATTCCAA	GTATAATGCA
1210	1220	1230	1240	1250	1260
GAAAGCACAG	AAAGGGAGTC	ACAAGACACT	GTTCGACAGA	ATGATGATGG	CGGGTTTCACT
1270	1280	1290	1300	1310	1320
GAGGAATGSG	AAGCCGACAG	GGACAGTCAT	GTAGGGGCTC	ATCGCTCTAC	ACCTGAGTCA
1330	1340	1350	1360	1370	1380
CGAGCTGCTG	TCCAGGAAC	TTCCAGCAGT	ATCTCTGCTG	GTGAAGACCC	AGAGGAAAGG
1390	1400	1410	1420	1430	1440
GGAGTAAAC	TTGGATTGGG	AGATTTCATT	TTCTACAGTG	TTCTGGTTTG	TAAAGCTTCA
1450	1460	1470	1480	1490	1500
GCAACAGCCA	GTGGAGACTG	GAACACAACC	ATAGCTCTTT	TCCTAGCCAT	ATTAATTGGT
1510	1520	1530	1540	1550	1560
TTGTGCTTCA	CATTATTAAT	CTTGTCCATT	TTCAAGAAAG	CATTGCCAGC	TCTTCCAAAT
1570	1580	1590	1600	1610	1620
TCCATCACCT	TTGGGGTTGT	TTTCTACTTT	GCCACAGATT	ATCTTGATCA	GCCTTTTATG
1630	1640	1650	1660	1670	1680
GACCAACTAG	CATTCCATCA	ATTTTATATC	TAGCATATTT	GCGGTTAGAA	TCCCATGGAT
1690	1700	1710	1720	1730	1740
GTTCCTTCTT	TGACTATAAC	CAATCTGGG	GAGGACAAAG	GTGATTTTCC	TGTGTCCACA
1750	1760	1770	1780	1790	1800
TCTAACAAAG	TCAAGATTCC	CGGCTGGACT	TTTGCAGCTT	CCTTCCAGT	CTTCCTGACC
1810	1820	1830	1840	1850	1860
ACCTTGCACT	ATTGGACTTT	GGAAGGAGGT	GCCTATAGAA	AACGATTTTG	AACATACTTC
1870	1880	1890	1900	1910	1920
ATCGCAGTGG	ACTGTGTCTT	CGGTGCAGAA	ACTACCAGAT	TTGAGGGACG	AGGTCAAGGA
1930	1940	1950	1960	1970	1980
GATATGATAG	GCCCCGAAGT	TGCTGTGCCC	CATCAGCAGC	TTGACGCGTG	GTCACAGGAC
1990	2000	2010	2020	2030	2040
GATTTCACCTG	ACACTGCGAA	CTCTCAGGAC	TACCGGTTAC	CAAGAGGTTA	GGTGAAGTGG
2050	2060	2070	2080	2090	2100
TTTAAACCAA	ACGGAACCTT	TCATCTTAAA	CTACACGTTG	AAAATCAACC	CAATAATTCT
2110	2120	2130	2140	2150	2160
GTATTAACCTG	AATTCTGAAC	TTTTCAGGAG	GTACTGTGAG	GAAGAGCAGG	CACCAGCAGC

2170	2180	2190	2200	2210	2220
ACAATGGGGA	ATCGAGAGGT	GGGCAGGGGT	TGCAGCTTCC	CTTCAATTTT	TTGCTGCAGA
2230	2240	2250	2260	2270	2280
CTCATCTTTT	TAAATGAGA	CTTGTCTTCC	CTCTCTTTTG	AGTCAAGTCA	AATATGTAGA
2290	2300	2310	2320	2330	2340
TGCCTTTGGC	AATCTTTCTT	CTCAAGCACT	GACACTCATT	ACCCTCTGTG	ATTGCCATTT
2350	2360	2370	2380	2390	2400
CTTCCCAAGG	CCAGTCTGAA	CCTGAGGTTG	CTTTATCTTA	AAAGTTTAA	CCTCAGGTTT
2410	2420	2430	2440	2450	2460
CAAATTCAGT	AAATTTTGGG	AACAGTACAG	CTATTTCTCA	TCAATTCCTT	ATCATGTTGA
2470	2480	2490	2500	2510	2520
AGTCAAAATTT	GGATTTTCCA	CCAAATCTTG	AATTTGTAGA	CATACCTGTA	CGCTCACTTG
2530	2540	2550	2560	2570	2580
CCCCAGATGC	CTCCTCTGTC	CTCATTCTTC	TCTCCACAC	AAGCAGTCTT	TTTCTACAGC
2590	2600	2610	2620	2630	2640
CAGTAAGGCA	GCTCTGTGCT	GGTAGCAGAT	GCTCCCACTT	ATTCTAGGGT	CTTACTCTTT
2650	2660	2670	2680	2690	2700
GTATGATGAA	AAGAATGTGT	TATGAATCGG	TGCTGTCAGC	CCTGCTGTCA	GACCTCTCTT
2710	2720	2730	2740	2750	2760
CACAGCAAT	GAGATGTATG	CCCAAAGCGG	TAGAAATAAA	GAAGAGTAAA	ATGGCTGTTC
2770	2780	2790	2800	2810	2820
AAGCAAAAAA	AAAAAAAAAA	AAAAAAAAAA	A.....

== Protein

SEQUENCE ID NO: 2

MTFLPAPLSYFQNAQMSZDNHLSNTVRSQNDNRERQEHNDRRSIGZPEPLSNGRPQGNRQVVEQ
DEEEDDELTLYGAKGVIMLFVPVTLGVVVVATIKSVSPYTRKDGQLITPFTEDTETVGQRAL
HSILNAAIMISVIVVMTILLVVLXKRCYKVIHAWLIISLILLFFFSFIYLGVEVFKTYNVXVDY
ITVXLLIWNLGUVGMISIHWKGFRLXQAYLIMISALMALVFIKYLPEWTAWLILAVISVYDLVA
VXCPKGPLRMLVETAQERNETLFPALYSSSTVWLVMNAEGDPEAQRVSKYKYNASTERESQ
DTVAENDCGGFSEWEAQORDSHLGPHRSTPESRAAVQELSSSILAGEDPEERGVLGLGDFIFYS
VLVGKASATASGDWNTTIACFVAILIGLCITLILLAIFKKALPALPISITFGLVIFYFATDYLVPF
FMDQLAFHQFYI

DNASTS

BASES .000 .000

Sequence ID NO 3

10	20	30	40	50	60
ACCTACACAC	GGGCGCTTAC	GGGGAACCT	AGGCTGGAG	GGGCGCTTAC	GGGCGCTTAC
70	80	90	100	110	120
AGAGAGGGA	CGACACAG	ACAGCAGCT	TTGCTGCTT	TTAGCGGCT	TGAGGAGGA
130	140	150	160	170	180
CACATGAGC	AAAGATGCT	AAGAGCTTT	GTCTTCTTC	AGAAGGTAT	TCTTCCAGC
190	200	210	220	230	240
TCTTCTTAT	ACAGAGATC	CTGACCTTT	GTCTTCTTC	CAGATGCTT	AGAGCTTGA
250	260	270	280	290	300
GGACAGGAC	TGAGGAGG	CGATCGGAG	CGAGATGAC	AGCTACAC	GGGAGGGA
310	320	330	340	350	360
CGATGAGG	CAGAGCTG	ACAGCTTGA	CGATATCT	AATCGGCTT	CGGAGGTA
370	380	390	400	410	420
CTGAGGAG	GTCTTCTTC	AAGATGAG	CGAGAGGA	GAGCTGCTT	TGAAATGCT
430	440	450	460	470	480
AGGAGGCT	GTCTTCTTC	TCTTCTTCT	CGGAGCTT	TGATGCTT	TCTTCTTCT
490	500	510	520	530	540
CGGATGGA	TGATGCTT	TGATGCTT	GAGGAGCT	CAGCTATCT	ACAGGCTT
550	560	570	580	590	600
CGGAGGAG	AGGAGCTT	TGAGGAGG	AGGCTGCT	TGATGCTT	ATGCTGCTT
610	620	630	640	650	660
CGGATGCT	GTCTTCTTC	TGATGCTT	CGGCTGCT	GTCTTCTTC	AGGAGGCT
670	680	690	700	710	720
CGGAGGCT	ATGCTGCTT	GGGATGCT	TGATGCTT	TGCTGCTT	TCTTCTTCT
730	740	750	760	770	780
GTCTATTC	TGAGGAGG	TGATGAGC	CGGATGCT	AGGCTGCT	AGCTGAGT
790	800	810	820	830	840
AGGATGCT	ATGCTGCT	GGGCTGCT	CGGAGGCT	CGGATGCT	CGGAGGCT
850	860	870	880	890	900
CTCTGAGT	CAGGAGCT	ATCTATTC	GAGGAGCT	CTCTGAGT	TGATATTC
910	920	930	940	950	960
CAATGAGT	CGGAGGCT	CGGATGCT	CGCTGCTT	GTCTTCTTC	TGATGAGT
970	980	990	1000	1010	1020
GTCTGAGT	TGATGCTT	AGGCTGCT	TGATGCTT	GTCTGAGG	CTCTGAGG
1030	1040	1050	1060	1070	1080
AAATGAGT	CTCTTCTTC	CTCTTCTTC	TGCTGAGG	ATCTGCTT	TGCTGAGT
1090	1100	1110	1120	1130	1140
CGCTGAGG	GAGGAGG	CGGAGGAG	GTCTGAGG	AGGCTGCT	ATCTGAGG
1150	1160	1170	1180	1190	1200
AGGCTGAG	AGGAGGAG	AGGAGGCT	TGCTGAGG	GATCTGCT	CTCTGAGG
1210	1220	1230	1240	1250	1260
CGCTGAGG	CGGAGGAG	AGGAGGCT	CGGCTGCT	CGCTGAGT	CTCTGAGG

DNASIS ~~Seq~~
~~data~~

GN					
1270	1280	1290	1300	1310	1320
AGCTGCTTC	CAGGAACTT	CTGGAGATC	TCTAACGAT	GAAGCTGCT	AGGAAAGGG
1330	1340	1350	1360	1370	1380
AGTAAACTT	GGATGGGAC	ATTGATTTT	CTGAGTCTT	CTGTTTGGA	AGGCTTCAG
1390	1400	1410	1420	1430	1440
AACGGGAGT	GGAGCTGCA	ACACACCAT	AGCTGCTTC	GTAGGCTAC	TGATGGGCT
1450	1460	1470	1480	1490	1500
CTGCTTAAA	TGATCTGTC	TGCTCTTAA	CAGAAAGG	TAGGAGGCT	AGGCTATCT
1510	1520	1530	1540	1550	1560
CAGGCTGTC	GGGTTGCTG	TGCTCTTGC	CAGGCTTAC	CTTGGGAGC	CTTCTAGGA
1570	1580	1590	1600	1610	1620
CGAAGTCCA	TGAGTCACT	TTTACATCA	GGCTTTGTC	AGTTAGACA	TGGATCTTC
1630	1640	1650	1660	1670	1680
TTCTTTGAT	ATTAAACCA	CAAAACAGA	GAGGAGGCT	GAGGAGAGA	CTGTTGCTT
1690	1700	1710	1720	1730	1740
TTCTTTGTC	TGAGTCACT	AAGGAGGAC	TGAGGCTGA	CTTGGGAGC	TTCTTTGGA
1750	1760	1770	1780	1790	1800
CTTGGGAGC	CGAAGTCCA	TGCTCTTGC	TGAGGAGAG	CTTGGGAGA	GGAGGCTTT
1810	1820	1830	1840	1850	1860
CGAAGTCCA	TGCTCTTGC	AGAGGCTTC	CTTGGGAGC	TGAGGAGCA	AGGAGAGGA
1870	1880	1890	1900	1910	1920
AATGCTGGG	GGAGGAGAG	TGCTCTTGC	TGAGGCTTC	GGAGGCTTC	CTTGGGAGC
1930	1940	1950	1960	1970	1980
TAGGCTGAC

ONASIS ~~h~~
~~DATE: 0000000000~~

SEQUENCE ID NO 4

20	20	30	40	50	60
WTEPAPLEY	FORQOESDS	HESSEALRSCN	DSCEHQQCHD	RCRLNFPPT	SNCRPQSNR
70	80	90	100	110	120
QVZQZED	ZLLTETKX	HVEMLFVT	LQWVWAT	KEVEPTKED	QLLTPTPS
130	140	150	160	170	180
QSTVQORAL	HSTLALMT	SVTVMFLL	WLYKRCYK	VEHAWLISS	LILLPTFTI
190	200	210	220	230	240
YLGVTFTN	WVDTYVAL	LZANGVGN	DAENKPLR	LQAYLMS	ALMALVFT
250	260	270	280	290	300
LPENTAWLL	AVISVYELVA	VLCPKFLM	LVEDQERNR	TLFPALYSS	TRVWUTRAE
310	320	330	340	350	360
QOPEACRV7	KMPKNTQRA	ZRTQSGSG	NDOGZSEW	ZACRUSLGT	HSTPESRAA
370	380	390	400	410	420
VQELSGSST	SEDFEERGK	LQWTFPYS	VLVGRATA	SGRNTTAC	XVAILLGL
430	440	450	460	470	480
XLLATKX	QOAPLST	FGVFAFAD	YLVQFMQ	AFQYF...

SEQUENCE ID NO: 5

```

10      20      30      40      50      60
GAATTCGGCA GAGGTAAT GTGTTTCTT CAGAGACTC TCAGGGGCA GTTGGCTTGG

70      80      90      100     110     120
GCCCGGATTA GTAGCTCTCT GAACTGGAGT GGAGTAGGAG AAAGAGGAAAG CGTCTTGGGG

130     140     150     160     170     180
TGGGTCTGCT TCAGCAACTG GTGAAACTCC GGTCTCAGC CCGCGGGTGT GTCTTTGTCC

190     200     210     220     230     240
AGGGGGCAGC AGCATTCTGG GCGAAGTCCG CACGCTCTTT GTTCGAGGCC GAAGACGGGG

250     260     270     280     290     300
TCTGATCTTT TCTGCTTGGT CGGACTGTCT TCAGGGCATG CATCTCCAGT GACTCTTTGT

310     320     330     340     350     360
TTTGCTGCTG CTTCCTCTCT AGATTCTTCT CACCTTTGTC GTGAGCTCTG CTTTAGGCA?

370     380     390     400     410     420
TATTAATCCA TAGTGGAGGC TGGGATGGGT GAGAGAATTG AGGTGACTTT TCCATAATTC

430     440     450     460     470     480
AGACCTAATC TGGGAGCCTG CAAGTCACAA CAGCTTTGCC GGTCTTTAGA CAGCTTGGCC

490     500     510     520     530     540
TGGAGGAGAA CACATGAAGG AAAGAACCTC AAGAGGCTTT GTTTTCTGTG AAACAGTATT

550     560     570     580     590     600
TCTATACAGT TGCTCAATG ACAGAGTTAC CTGCACCTTT GTCTTACTTC CAGAATGCAC

610     620     630     640     650     660
AGATGTCTGA GGACAACCAC CTCAGCAATA CTATGACAA TAGAGAACCG CAGGAGCACA

670     680     690     700     710     720
ACGACAGACG GAGCCTTGGC CACCTGAGC CATTATCTAA TGGACGACCC CAGGGTAACT

730     740     750     760     770     780
CCCGGCAGGT GGTGGAGCAA GATGAGGAAG AAGATGAGGA GCTGACATTG AAATATGGCG

790     800     810     820     830     840
CCAAGCATGT GATCATGCTC TTGTCTCTG TGACTCTCTG CATGGTGGTG GTCGTGGCTA

850     860     870     880     890     900
CCATTAAGTC AGTCAGCTTT TATACCCGGA AGGATGGGCA GCTAATCTAT ACCGCATTCA

910     920     930     940     950     960
CAGAAGATAC CGAGACTGTG GGCCAGAGAG CCCTGCACTC AATTCTGAAT GCTGCCATCA

970     980     990     1000    1010    1020
TGATCAGTGT CATGTTGTCT ATGACTATCC TCCTGGTGGT TCTGTATAAA TACAGGTGCT

1030    1040    1050    1060    1070    1080
ATAAGGTCAT CCATGCCCTGG CTTATTATAT CATCTCTATT GTTCTCTTTC TTTTCTTCAT

```

1090	1100	1110	1120	1130	1140
TCATTTACTT	GGGGGAAGTG	TTTAAAGCCT	ATACCTTTCC	TGTCGACTAC	ATTACTTTTG
1150	1160	1170	1180	1190	1200
CACCTCTGAT	CTGGATTTTG	GGTGTGGTGG	GAATCATTTG	CATTGACTCG	AAAGGTTGAC
1210	1220	1230	1240	1250	1260
TTGCACTGCA	CCAGCCATAT	CTCAATTATG	TTAGTGGCCT	CATGGCCCTG	GTGTTTATCA
1270	1280	1290	1300	1310	1320
AGTACCTCCC	TGAATGGACT	GGGTGGCTCA	TCTTGGCTGT	GATTTCAGTA	TATGATTTAG
1330	1340	1350	1360	1370	1380
TGGCTGTTTT	GTGTCCGAAA	GGTCCACTTC	GTATGCTGGT	TGAAACAGCT	CAGCAGAGAA
1390	1400	1410	1420	1430	1440
ATGAACGGCT	TTTTCCAGCT	CTCATTTACT	CTCAACAAT	GGTGTGGTTG	GTGAATATGG
1450	1460	1470	1480	1490	1500
CAGAAGGAGA	CCCGGAAGCT	CAAAGGACAG	TATCCAAAA	TTCCAGTAT	AATCCAGAAA
1510	1520	1530	1540	1550	1560
GCACAGAAAG	GGAGTCACAA	GACACTGTTG	CAGGCAATGA	TGATGGCCGG	TTCACTGAGG
1570	1580	1590	1600	1610	1620
AATGGGAAGC	CCAGACGGAC	AGTCATCTAG	GGCTTCATCG	CTCTACGCTT	GAGTCACCTAG
1630	1640	1650	1660	1670	1680
CTGCTCTCCA	GGAACTTTCC	AGCAGTATCC	TGGCTGGTGA	AGACCCAGAG	GAAAGGGGAG
1690	1700	1710	1720	1730	1740
TAAACTTTGG	ATTGGGAGAT	TTCAATTTCT	ACAGTGTTCCT	GGTTGGTAAA	GCCTCAGCAA
1750	1760	1770	1780	1790	1800
CAGCCAGTGG	AGACTGGAAC	ACAACCATAG	CCTGTTTCCT	AGCCATATTA	ATTGGTTTGT
1810	1820	1830	1840	1850	1860
GCCTTACATT	ATTACTCTTT	GCCATTTTCA	AGAAAGCAAT	GCCAGCTCTT	CCAACTCTCA
1870	1880	1890	1900	1910	1920
TCACCTTTGG	GCTTGTTTTC	TACTTTGCCA	CAGATTATCT	TGTACAGCCT	TTTATGGACC
1930	1940	1950	1960	1970	1980
AATTAGCATT	CCATCAATTT	TATATCTAGC	ATATTGGGGG	TTAGAAATCC	ATGGATTTTT
1990	2000	2010	2020	2030	2040
CTTCTTTGAC	TATAACCAAA	TCTGGGGAGG	ACAAAGGTGA	TTTTCTGTG	TCCACATCTA
2050	2060	2070	2080	2090	2100
ACAAAGTCAA	GATTCCCGGC	TGGACTTTTG	CAGCTTCCTT	CCAAGTCTTC	CTGACCAGCT
2110	2120	2130	2140	2150	2160
TGCACATTTG	GACTTTGGAA	GGAGGTGCTT	ATAGAAACG	ATTTTGAACA	TACTTCATCG

2170	2180	2190	2200	2210	2220
CAGTCGACTG	TGTCCTCGGT	GCAGAAACTA	GCAGATTTTA	GGGACGAGGT	CAAGGAGATA
2230	2240	2250	2260	2270	2280
TGATAGGCCC	GGAAAGTTGCT	GTGCCCCATC	AGCAGTTTGA	CGCCTGGTCA	CAGGACGATT
2290	2300	2310	2320	2330	2340
TCAGTGACAC	TGCGAAGTGT	CAGGACTACC	GGTTACCAAG	AGGTTAGGTG	AAGTGGTTTA
2350	2360	2370	2380	2390	2400
AACCAACGG	AAGTCTTCAT	CTTAAACTAC	ACGTTGAAAA	TCAACCCAAT	AATTCTGTAT
2410	2420	2430	2440	2450	2460
TAACTGAATT	CTGAAGTTTT	CAGGAGGTAC	TGTGAGGAAG	AGCAGGCACC	AGCAGCAGAA
2470	2480	2490	2500	2510	2520
TGGGGAATGG	AGAGGTGGGC	AGGGGTTCGA	GCTTCCCTTT	GATTTTTTGC	TGCAGACTCA
2530	2540	2550	2560	2570	2580
TGCTTTTAAA	ATGAGACTTG	TTTTCCCTTC	TCTTTGAGTC	AAGTCAAATA	TGTAGATGCC
2590	2600	2610	2620	2630	2640
TTTGGGAATT	CTTCTTCTCA	AGCACTGACA	CTCATTACCG	TCTGTGATTG	CCATTCTCTC
2650	2660	2670	2680	2690	2700
CCAAGGCCAG	TGTGAACCTG	AGGTTGCTTT	ATCCTAAAAG	TTTTAACCTC	AGGTTCCAAA
2710	2720	2730	2740	2750	2760
TTGAGTAAAT	TTTGGAAACA	GTACAGCTAT	TTCTCATCAA	TTCTCTATCA	TGTTGAAGTC
2770	2780	2790	2800	2810	2820
AAATTGGGAT	TTTCCACCAA	ATTCTGAATT	TGTAGACATA	CTTGTACGCT	CAGTTGCCCC
2830	2840	2850	2860	2870	2880
AGATGCCCTC	TCTGTCTCTA	TTCTTCTCTC	CCACACAAGC	AGTCTTTTTC	TACAGCCAGT
2890	2900	2910	2920	2930	2940
AAGGCAGCTC	TGTGCTGGTA	GCAGATGGTC	CCACTTATTC	TAGGCTCTTA	CTCTTTGTAT
2950	2960	2970	2980	2990	3000
GATGAAAAGA	ATGTGTTATG	AATCGGTGCT	GTCAGCCCTG	CTGTCAGACC	TTCTTCCACA
3010	3020	3030	3040	3050	3060
GCAATGAGA	TGTATGCCCA	AAGCGGTAGA	ATTAAAGAAG	AGTAAAATGG	CTGTTGAAGC
3070	3080	3090	3100	3110	3120
AAAAAAAAAA	AAAAAAAAAA	AAAAAA...

SEQUENCE IN NO: 6

843-397.697

10	20	30	40	50	60
GTCTTCCTAA	CCAACTTAGG	AGNTTCGACC	TGGGAAGAC	CNACNTGATC	TCCCGGAGCN
70	80	90	100	110	120
AAAGACTTCA	GTTGAGCCCT	GATTGCACCC	ACTTTACTCC	AAGCCTGGCC	AACCAAAATG
130	140	150	160	170	180
AGACACTGGC	TCCAAACACA	AAAACAAAA	CAAAAAAGA	GTAAATTAA	TTANAGCGAA
190	200	210	220	230	240
CNATTAATA	AATAATAGCA	CAGTTGATAT	AGGTTATGGT	AAAATTATA	AGGTGGGANA
250	260	270	280	290	300
TTAAATCTTA	ATGTTTGGCA	GCCATCAGAT	TATCTAAAT	AATGTTTTCG	TGGAAATTA
310	320	330	340	350	360
TGTACATCTT	TTAAATCTG	TGTAATTTT	TTTCAGGGAA	GTGTTTAAA	CTATATACCT
370	380	390	400	410	420
TGCTCTGGAC	TACATTAATG	TTCACCTCTT	GATCTGGAAT	TTTCTGTCG	TGGGAATGAT
430	440	450	460	470	480
TTCCATTAC	TGGAAAGGTC	CAGTTCCACT	CCAGCAGCCA	TATCTCATTA	TGATTAGTGC
490	500	510	520	530	540
CTTCATGICC	CTGKATTTA	TCAAGTACCT	CCCTGAATGG	ACTGNGTGGC	TCACTTTGGC
550	560	570	580	590	600
TGTGATTTCA	GTATATGCTA	AAACCAAGA	CTCATTAATT	GTCTCTCACA	GCAATGCCCC
610	620	630	640	650	660
ACTGCACTGT	TTCTTTTCTT	CATCTCTTTA	TCTTGATTTA	GAGAAATGCG	TAAGCTGTAC
670	680	690	700	710	720
ATCCCATTAAC	TCTTCAGTAA	ATCATTAATT	AGCTATAGTA	ACTTTTTCAT	TTGAAGATT
730	740	750	760	770	780
CGGCTGGGCA	TGGTAGCTCA	TGGCTGTAAT	CTTAGCACTT	TGGGAGGCTG	AGCCCGGCGC
790	800	810	820	830	840
ATCACCTAAG	CCCAAGACTTC	AAGACCAGCC	TGGGCAACAT	GGCAAAACCT	CGTATCTACA
850	860	870	880	890	900
GAAAATACAA	AAATTAGCCG	GGCATGGTGG	TGCACACCTG	TAGTTCGAGC	TACTTAGGAG
910	920	930	940	950	960
GCTGAGGTCG	GAGGATCGAT	TGATCCCAAG	AGGTCAGGNC	TGCAG.....

SEQUENCE IN NO: 7

874-125m.125

10	20	30	40	50	60
GTTCGAAGT	CATGGATTCC	TTTAGGTAGC	TACATTATCA	ACCTTTTTCG	GAATAAAATG
70	80	90	100	110	120
AATTGAGAGT	GTTACAGTCT	AATTCTATAT	CACATGTAAC	TTTATTTTGG	ATATATCAGT
130	140	150	160	170	180
AAATAGTGGT	TTTTTTTTTT	TTTTTTTTTT	TTTTTTTTTT	TTTGGGGGANA	GAGTCTCGCT
190	200	210	220	230	240
CTGTCCCCAG	GTTGGAGTGC	AATGGTGGCA	TCTTGGCTCA	GTGAAGGCTC	CACCNCCCGG
250	260	270	280	290	300
GTTCAAGTGA	TTTCTCTGCC	TCAGCCNCCC	AAGTAGNTGG	GACTACAGGG	GTGCGCCACC
310	320	330	340	350	360
ACCCCTCGGA	TAATTTTGGG	NTTTTTAGTA	GAGATGGGCT	TTCACCAGCT	TGGNCCAGGG
370	380	390	400	410	420
TGGTCTTCCA	ACTCCTGANA	TCATCATCTG	CCTCCCTTAG	CCTCCCCAAA	GTGCTGGGAT
430	440	450	460	470	480
TTCAGGGGTC	AGCTACTGTT	CCTCGGGGTC

SEQUENCE DA NO: 8

886-1000.100

10	20	30	40	50	60
GCTCATCATG	CTTCACGCGG	GAGGCTGTGC	GCGAAGAATC	CTCCGACACA	GNATTAAGAA
70	80	90	100	110	120
TGCTCCGCCA	CAGGATAGAG	AATGCGGCGG	CACAGCATAG	AGAACGCGCG	GCACAGCATA
130	140	150	160	170	180
GAGAATGCGG	CCGACACAGCA	TAGAGAAGCC	CCGCGACAGC	ATAGAGAATG	CTCTTCACCT
190	200	210	220	230	240
CTGGGTTTTT	AACCGCGGAA	ACTAAATATC	CAGAGGCGCA	CACATCATTT	AAGATAGAAA
250	260	270	280	290	300
TTCTCGATC	TTTTAAATTT	TTTCMAAGTA	GTTTTACTTA	TTTTGAGATT	CTATTTCTTT
310	320	330	340	350	360
ACTAGCAATTA	AGGGATAAAA	TAACAATGTC	TGCATAATGA	ACCGTATGAA	ACMAACGMAA
370	380	390	400	410	420
GCTAGGTTTT	TTTCATAGST	CTTCTTCGAG	ATTGAATGAA	CGTGTGTTCT	AAAATTTAAC
430	440	450	460	470	480
CCCCCAGCGA	AATATTCAGT	TAACTATGTT	AAAAACCGAG	ACTTGTGATT	GAGTTTTGCC
490	500	510	520	530	540
TGAAAATGCT	TTCAATATTA	TGTGTGAATG	TGTGTG....

SEQUENCE NO: 9

901-312.988

10	20	30	40	50	60
GGATCCCTCC	CTTTTTCAGA	CCATACAAGG	TAACTTCCGG	ACCTTGCCAT	GGCATCTGTA
70	80	90	100	110	120
AACCTGTCATG	GTCTTCCCGG	GGAGTCTCTT	TAGCATGCT	AATGTATTAT	AACCTAGCTTA
130	140	150	160	170	180
TAGTGAGCAG	TGAGGATAAC	CAGAGGTCAC	TCTCTTCACC	ATCTTGGTTT	TCTCTGGTTT
190	200	210	220	230	240
TGCTCAGCTT	CTTTATTGCA	ACCACTTTTA	TGAGCAAGAT	CTTTATGAGC	TGTATCTGT
250	260	270	280	290	300
GCTGACTTCC	TATCTCATCC	CGTAACCTAG	AGTACCTAAC	CTCTTCCAAA	TTGAGAGCCA
310	320	330	340	350	360
GNAGGCTCTTG	GNCTTATTTC	ACCCAGCCCG	TATTCAARAT	AGAGTNGTTC	TTGAGCCAAA
370	380	390	400	410	420
CGCCCTCAGC	ACAAGGATTT	TAAAGTCTTA	TAAATTAAAG	TAAAGTAGAT	CGTTGATAT
430	440	450	460	470	480
GTGCTCTGAA	ATCACAGAAA	GCTGAACTTC	GAAGAAAGGT	CTTGGAGCTC	CAGCCAGTAA
490	500	510	520	530	540
ACAAGTCTTC	ATGCAAGGTT	CAGTATTTAA	GCTACATCTC	AAAGGATTAAG	TACATTTCTC
550	560	570	580	590	600
TATCTTGGCA	TGAACAGAGA	GAATGGAGCA	ATCCAGAGCC	CAGCTAAAG	AGAGGAGCTC
610	620	630	640	650	660
AATGCTCTCA	GTGAACATG	ATAGATTAAT	TAGACTTTTA	AAGTGCATAC	TTCTCTTACA
670	680	690	700	710	720
TTCTTTTCTC	TTGCTTCAGG	TTTTTAGAAC	TCTATGTCAC	GGTCTCTTTC	TTAATCCAG
730	740	750	760	770	780
GTCTAACCGT	TACCTTGATT	CTGCTGAGAA	TCTGATTTAC	TGAAATGTT	TTCTCTTCTC
790	800	810	820	830	840
TTATAGAAATG	ACAATAGAGA	ACCGCAGGAG	CACAAAGACA	GACGGAGCCT	TGCCCACCTT
850	860	870	880	890	900
CANCCATTAT	CTAATCGAGC	ACCCAGGGTA	ACTCCCGGCA	GGTGGTGGAN	CAGATGAGG
910	920	930	940	950	960
AAGAAGATGA	GGANCTGACA	TTGAAATATG	NCGSCAAGCA	TGTGATCATG	CTCTTTGACC
970	980	990	1000	1010	1020
CTGTGACTCT	CTGCATGGTC	GTCTCTGTCG	HTACCATTA	GTCACTCAGC	TTTTATACCC
1030	1040	1050	1060	1070	1080
GGAGGAGATGG	GCAGCTGTAC	GTATGAGTTT	KGTTTTATTA	TTCTCAAASC	CAGTGTGGCT
1090	1100	1110	1120	1130	1140
TTCTTTTACA	GCAATGTCATC	ATCACCTTGA	AGGCTCTTNC	ATTGAAGGGG	CATGACTTAC
1150	1160	1170	1180	1190	1200
CTGGAGAGCC	CATCTCTCTCT	CATCTCTCAG	AGCAGTTGAG	AGANCCAGGG	GTTATTACTT
1210	1220	1230	1240	1250	1260

CATCTTTTAA	GTGGAGAAAA	GGAACACTGC	AGAAGTATCT	TTCCTGTATG	GTATTACTGG
1270	1280	1290	1300	1310	1320
ATAGGGGTGA	AGTTATGCTG	AATTGAACAC	ATAAATTCTT	TTCGACCTCA	GGGNCATTGG
1330	1340	1350	1360	1370	1380
GGGGCCATTG	NTCTTCTGCC	TAGAATATTC	TTTCCTTTTC	TNACTTGGGN	GGATTAAATT
1390	1400	1410	1420	1430	1440
CCCTGTGATCC	CCCTCTCTCT	GGTCTTATAT	ATAAAGTNTT	GGTGCCCTCA	AAGAAGTAGC
1450	1460	1470	1480	1490	1500
ACTCGAATAT	AAAATTTTTC	TTTTAAATCT	CAGCAAGGNA	AGTTACTTCT	ATATAGAAGC
1510	1520	1530	1540	1550	1560
GTGCACCCNT	ACAGATGGAA	CAATGGCAAG	CSCACATTTG	GGACAAGGGA	GGGAAAAGGG
1570	1580	1590	1600	1610	1620
TTCTTATCCC	TGACACACCT	GGTCCCTGCT	GNTGTGTNCT	NCCCTTACTG	ANTAGGGTTA
1630	1640	1650	1660	1670	1680
GACTGGACAG	GCTTAACTTA	ATTCCTAATG	GNTAATTTRA	AGAGAATNAT	GGGTTGAATG
1690	1700	1710	1720	1730	1740
CTTTGGGAGG	AGTCRAGGAA	GAGNAGGTTAG	NAGGTAACCT	GAATCA....

SEQUENCE ID NO: 10

310-913.098

10	20	30	40	50	60
CNCSTATAAA	AGACCAACAT	TGCCANCYAC	AACCAACGGC	AAGATCTTCT	CTTACCTTCC
70	80	90	100	110	120
CCCYNGGCTGT	AATACCAAGT	ATTGNCCTAA	TTGTGATAAA	CTTTCATTGG	AAAGTGACCA
130	140	150	160	170	180
CCCTTCCTTCC	TTAATACATT	GTCTGTGCTT	GCTTTCACAC	TACAGTAGCA	CAGTTGAGTG
190	200	210	220	230	240
TTTCCCTCTCC	AGACCAATAG	ACCCATAGAG	CTTAAATAT	TCAGTCTGGC	TTTTTACAGA
250	260	270	280	290	300
GATGTTTCTG	ACTTTGTTAA	TAGAAAATCA	ACCCAACTGG	TTTAAATTAAT	GCACATACCT
310	320	330	340	350	360
TCTCTCTCAT	AGAGTAGTGC	AGAGGTAGNC	AGTCCAGATT	AGTASGGTGG	CTTCACCTTC
370	380	390	400	410	420
ATCCAAAGGAC	TCAATCTGCT	TCTTTCTTCT	TTAGCTTTCT	ACCTCTAGCT	TACTTCAGGG
430	440	450	460	470	480
TCCAGGCTGG	AGCCCTATCC	TTCAATTTCT	ACAGTAGGAA	GGAGTAGGGG	AGAAAAGAAC
490	500	510	520	530	540
ATAGGACATG	TCAGGAGAAT	TCTCTCTCTA	GAGGTTCTAT	ACCAACACA	TCTCCCTAGA
550	560	570	580	590	600
AGTCAATGCC	CTTACTTCTT	CTCAATAGCA	TCTTAAATAT	AAGGGAGTCA	GAAGTAAAGT
610	620	630	640	650	660
CTXXNTGGCT	GGGAATATTC	GCACCTGGAA	TAAAAATGTT	TTTCTCTGAA	TCAGAAACAA
670	680	690	700	710	720
GGGGAAGATG	GATATGTGAC	ATTATCTTAA	GACCACTCCA	GTTGCAATTA	CTCTGCAGAT
730	740	750	760	770	780
GAGAGGCACT	AATTATAAGC	CATATTAAGT	TTCTTCTGAC	AACCACTTGT	CAGCCCTTCT
790	800	810	820	830	840
GCTTTCCTTG	GCAGAACTG	GTTCTATATC	AAGTTCTTAA	TAANCTGTAS	CCYAAAAAAT
850	860	870	880	890	900
TTCAATGAGT	ATTATAATTA	TTTCAATATA	AAGCACTCAC	TAGATGGAGC	CAGTGTCTGC
910	920	930	940	950	960
TTCAATGAGT	AAGTCTTCTT	TTCCATATGT	TAGACATTTT	CTTTGAAGCA	ATTTTAGAGT
970	980	990	1000	1010	1020
GTAGCTGTTT	TTCTCAGGTT	AAAAATTTT	AGCTAGGATT	GCTCAGTTCC	GGAAAGTGA
1030	1040	1050	1060	1070	1080
CTTATAAGAT	NCGAATGAA	TTAAGAAAAA	GAAAAATCTG	TGTTCCAGCT	GCTAATGCTG
1090	1100	1110	1120	1130	1140
KTGGTGATCT	YCATTAAACAC	TCANCTAGGG	CTTTXGXTT	TGXTTTATTC	TAGAACTAT
1150	1160	1170	1180	1190	1200
ACCCCACTCA	NAGAAGATAC	CGAGACTGTG	GGCCAGAGAG	CCCTGCACCT	AATTCCTAAT
1210	1220	1230	1240	1250	1260

GCTGCCATTA	TCATCAGAGT	CATTCTATCT	ATGACTAATNC	TCTGGCTGCT	TCATATATAA
1270	1280	1290	1300	1310	1320
TACAGGTGCT	ATAAGGTGAG	CATGAGACAC	AGATGTTTGH	TTTCCAGCTT	GTTCCTCTTA
1330	1340	1350	1360	1370	1380
TGGTTGGGTA	TTCTTTCTAC	AGTAACTTAA	CTGATCTAGG	AAAGAAAAAA	TGTTTTGCTT
1390	1400	1410	1420	1430	1440
TCTACAGATA	AGTAAATTTT	TAGTTTCTCT	CCTGCTCACT	GTCGAACATT	CAAAAAATAC
1450	1460	1470	1480	1490	1500
AAAAAGGAAG	CCAGGTGCTT	GTGTAATGCT	AGGCTCAGAG	GCTGAGGCTG	GAGGATGCTT
1510	1520	1530	1540	1550	1560
TGGGCTTAGG	AGTTCCACAAG	CAGCTTGGGC	AACCTAGCTA	GACCTGCTCT	CTATTAAACA
1570	1580	1590	1600	1610	1620
AAACAAAAAA	CAATATTTGG	AAGTATTTTA	TATGCACTGA	ATCTATATCT	CATGAAAAAA
1630	1640	1650	1660	1670	1680
TTAGTGTAAA	ATATATATAT	TATGATTAAG	TATCAAGATT	TATGATTAAT	TTATGTTATT
1690	1700	1710	1720	1730	1740
TTGGCATTTT	AATGCTTTT	TAGGCTATTC	TCTCAATATA	TAAAGCTGCA	AAACAAAAAA
1750	1760	1770	1780	1790	1800
AGTTGTAACT	CAAAAAATAA	CATTCTGATA	TAAATAGCTA	ATCTAAGTCT	GTCTTTTCTT
1810	1820	1830	1840	1850	1860
CTTTGTTTGH	TTCTTTGAAGC	AGGGCTCTTC	CCTTTCACCT	AGCTTCTGCT	CAAGTCTGCT
1870	1880	1890	1900	1910	1920
GGCAGCATTT	TGGCTCTCTC	CAG.....

SEQUENCE ID NO: 11

923-913.cen

10	20	30	40	50	60
CAGGAGTGGG	GTAGCTAAAT	GAGAGTCTT	TAAAGAGAG	ATGNGGNCNG	GGACATAGTG
70	80	90	100	110	120
GTACACAGCT	GTAAATGCTA	NCAGTATCG	GGAGTACTGA	AGGNGGNSGG	ATCACTTNG
130	140	150	160	170	180
GCTCTGGAAT	NTAGACACAG	CGTGGGCAAT	ATGGGCAATC	CGTCTCTCTA	GTAAAAATAG
190	200	210	220	230	240
CGAAATATTA	CGCTAGGCTG	GTGGGCGCTA	CGCTGGTTC	CACCTACTCA	GGAGGCTTAA
250	260	270	280	290	300
GCACGAGTAA	TCTTGAACG	CAGGAGGCG	AGGNTCTGGT	GAGCTGAGAT	CGTGGCTACT
310	320	330	340	350	360
CAGTCTGCTC	TGGGCACTA	AGTGGAGCTC	TCTCTCTTAA	AAGAAAAAAA	AAATCTCTAC
370	380	390	400	410	420
TTTTTAAGCG	TTGTGGGAGC	TCTTATATAT	ATTGAAATCG	TTCTTTTCTA	CGTCTGCTAT
430	440	450	460	470	480
GGCTGGGCTA	TTATATCTAT	TCTATCTTAA	CTCTCTCTCT	TTACATCTAT	TTACTTGGGG
490	500	510	520	530	540
TAACTTCTGA	AATTGGGGCT	CTCTCTCTCA	GAACTTAACT	CGTNGTGGCT	GTCTAGGCTAG
550	560	570	580	590	600
CATTTAAGCG	CAGCTACTTT	GTCTCTTAA	TACTCTGAG	TTTTTATCTT	TTCTAGATAT
610	620	630	640	650	660
AGCTCATACT	TGCTATCTAA	AAGACTAGNC	AGTATCTA	ATTGAGACTT	TCTCTCTTNG
670	680	690	700	710	720
CTCTCTGCTT	ATATATAGTA	GAACTGAAG	NACTTAAGA	CTACGCTTAA	TTCTAAGCCT
730	740	750	760	770	780
TTGGGGAGGG	ATTATATAGC	CTCTCTAGT	GAACTCTTGT	GCTATCTGAA	TCTTTNTTAA
790	800	810	820	830	840
GAAAGGGTNT	CAGGGAATNG	TATTAANACC	AAAAATATAT	GAT.....

SEQUENCE ID NO: 12

930-519.gem

10	20	30	40	50	60
GTCTTTCCCA	TCTTCTCCAC	AGAGTTTGTG	CTTACATCTA	TTACTCTCTG	CGATTTCCTA
70	80	90	100	110	120
GAAAGCATTC	TCAGCTCTTC	CAATCTCCAT	CACCTTTCCG	CTTCTTTTCT	ACTTTGCCAC
130	140	150	160	170	180
AGATTCTCTT	GTACAGGCTT	TTATGGACCA	ATTAGCATTC	CATCAATTTT	ATATCTAGCA
190	200	210	220	230	240
TATTTCCCGT	TAGAATCCCA	TGGATCTTTC	TTCTTTCACT	ATAACAAAAT	CTCGGCGGCA
250	260	270	280	290	300
CAAAGCTGAT	TTCTCTCTCT	CACATCTAAC	AAATCTAGAT	CTCGGCGCTC	ACTTTTGGAG
310	320	330	340	350	360
GTCTTTCCCA	AGTCTTCCCT	ACCACCTTCC	ACTATTCGAC	TTTCCAAAGCA	GGTCCCTATA
370	380	390	400	410	420
GAAAACGATT	TTCAACATAC	TTCTCTCCAG	TGGACTCTCT	CTCTGCTGCA	GAAACTACCA
430	440	450	460	470	480
GATTTCAGCG	ACGAGCTCCA	CGACATATCA	TGGGCGCGCA	AGTTCTCTCT	CGGCACTCGC
490	500	510	520	530	540
AGCTTCAGCC	GTGGTCACAG	GACGATTTTC	ACTGACACTC	CGACTCTCTA	GGACTACGCT
550	560	570	580	590	600
TACTAGAGAG	TTAGGTGAAG	TGTTTTAAAC	CAACCTAAC	TCTTCTCTTT	AAACTACAGC
610	620	630	640	650	660
TTGAAAATCA	ACCGAATTAAT	TCTGATTTAA	CTGATTTTCC	AACTTTTCAG	GAGGTACTCT
670	680	690	700	710	720
GAGGAAGAGC	AGGCTCCACC	ACCAGAAATC	CGAATCTAGA	GGTGGGCAAG	GGTTCCTAGT
730	740	750	760	770	780
TCCTTTTGAT	TTTTTC....

SEQUENCE ID NO: 13

932.C21

10	20	30	40	50	60
GGATCCCTCC	GGCTTGGGCT	CCCAAAAGTCC	TGGCATTTCA	GGCATCAGCC	ACCGCTGCTC
70	80	90	100	110	120
GCTGAGTCTC	CGATTTCCTC	CCAGCTCTAC	CGAGTTCTCT	CATCTTAAGC	AAGTCACTCA
130	140	150	160	170	180
ACTTCTCTCC	ATTGCTTTCC	CCTTNNGTAA	AATAAGNATC	TTATCTGNCC	NNCCTGCTTC
190	200	210	220	230	240
GGGCATTCTC	ATAAGGATTA	GATCAGATTA	TAGCATTTTC	CAAAATTAAA	AGCGCTAGAC
250	260	270	280	290	300
AAATCATTTT	ATGAAATATC	AAAGATTAGN	TTGATTTTCC	GGCAGCATAG	AAAAAGGAT
310	320	330	340	350	360
CTTCAAGACA	TTCTCTTAGC	GATTACTCAA	GCCTCCCTTT	TCTGKNWAA	TGCGANNCTC
370	380	390	400	410	420
ATTNAAKATC	CTTNTCTGGG	TTCAAAATCT	TTGCTTTCTT	CAGGCGCTTC	CTACTTATTC
430	440	450	460	470	480
CTAAGCAGTC	CTACCTTTAG	CTTATAGTAA	ATTCTCTAGT	TAGTTGAAAG	TCTGACAAA
490	500	510	520	530	540
TTAATACATT	CCTCGTTTAC	AAATCTCTCT	TATAGTATTT	TGATTTGENT	AAATGNATTT
550	560	570	580	590	600
ACTAGATTTT	AAGTAACAT	GGATCAGCTC	GTAATATCTT	ATTTCAGACC	TAATCTGGGA
610	620	630	640	650	660
GGCTTCAAGT	GACAACAGCC	TTTCCCTTCC	TACACAGCTT	TGCTCTGAC	GACAACAGCT
670	680	690	700	710	720
CAAAAGAAAG	TTTCTTTCTC	CTTAATCTAA	TCTATCGAAG	TCTTTTATAT	AACAGATATA
730	740	750	760	770	780
TTCTAGTCCA	CRAAGTTCTC	TTTTTCTTTC	GGTTTCTGTA	AGCTCAAGAG	GCTTTCTTTT
790	800	810	820	830	840
CGCTCAAAACA	GTATTCTTAT	ACAGTTCTCT	CGAATCTGAC	AGAGTTACCT	GCACCTCTCT
850	860	870	880	890	900
GCTCTTACTT	CCAGAAATCA	CAGATCTCTC	AGGACACCCA	CCTCAGCAAT	ACT.....

SEQUENCE ID NO: 14

PROGRAM blastn DATALIB nr Begin >1011.con

TCAGAAAATACTTNGGGCAATCAGGAATCAATGAGAACAGCTGATGCATAATTCCTCCTGTG
ATGGAATGTAATAGTAATTTAACAGTGTCTCTTTCTTTTAACTGCTCAAGGATACAGCAAATA
AAACAAAGCAATATGAAGCTGAGAAATAGGTATCAGATTATCATAAAAGTATAGATCAAAGG
AATCTGCTKCTNAGGTGGGCGAGCAGCCTCTAGAAGCSACNAGGGAGACTTTTAGAACTACCAT
TCTCCTCTATAAGTGGATCCNANGCCAGGAAACTTGATATTGAGNACAATGGCCTTACTGAAA
TAACCTGTGATCCACTCGGNTCATCATCTCCACCCACCATAAATTGATGAGTNCCTATAAT
ATTCCANCCAGNCGAAATACCTGGRAGGTTACTGAAAGGCNACNATCAGACGAAAATAAAGNATA
CCGTAGGTAAATTCTACAGT

Sequence ID NC: 15

PROGRAM blastn DATALIB nr Begin >1024.cen

GTCTCTAGATCTCTTCAAATTCATTNIGCGTATAGGAGCTGGGATTACCGCGGGTGCTGAA
CCAGACTTGCNCTCCAATGGATCCTCCANACNGGANGGGGGGTGGACTCACACCATTTCAGGGG
GCTCGTAAAGATCCTGTTTTGANTATNTNCGGTCAATTACNCCCCAA

SEQUENCE Pb No. 16

PROGRAM blastn DATALIB nr Begin >1024t.com

AATGTAACACCTAAACCTCAAACTCCTGNAACAATGTTACTTATNGATNCCATTNCTTTTT
NCACTCTCAGACATAAATATAAACCTANTTTCTACTGTGGRAAAACATCTN/CAGGGGNCNTTTAN
CCTATGATCTCTAGNACIANGGGCTNGTGGTNGTTTTAATGTCTCTAAGCTACTNGACTAGTTTC
TCTTNCACCTGAGNAACTGCAACAAGTNTTNCCTNCTGNATCTGNACTGNAATCCTAAGTTNCAA
GTNCCAATGAGCTNGTGANTTANCTTATTTTTAMCIIAAGTNTTAATCANCCNCAGTCTTACT
TTGNAAGCTNCTCCCTGGACAGGCTGCCCCNACTTCTAATGTTATGAATGGGCTGGAGNANCCTC
NACNTGAGTTTNNWAAGGNTCAACANCCATGNAANTGTAMCCGACTCTAAATTCCAACCNATA
AT

SEQUENCE ID NO:17

PROGRAM blastn DATALIB nr Begin >1058.con

ATCTGTGCTAGGTAGTGTACTAATCATTTCAGTTTATCTCATTTAATCTNNATGNAACTCTAAGTC
ATTGCTNTGANCNACACATAACAGATCTCCCACTGNAGTTTAGCGAGGCCAGTAAATTTXCCA
AAGNTCATAATVCTAAGNAGTTCTAGNATGGAGATTCTAAGTCCNACTGTTTACTCAAGAGACCC
TACTGTTAACTAGTACCTTTACACTACTAACTGGGTAANCCATAANCAATTAATGATAAAGATTG
AGATTACTKCCACATTCTCACTGGTTATAAATTAAAACNTCAAATAAAAANTCTTGGCACTTCTA
TGGTAATATTTTATTAGGATAAACTTTCAAGNACTGGATNCTAGGTG

SEQUENCE 2) NC:18

PROGRAM blastn DATALIB nr Begin >1092.con

CCCACACTGNTGGGCCATGGAAGCCATGAGTGTACCAACATGGCCCTGTCCCACTGGCCACAGTNG
ATTGGTTGNTCGGGAGTAGTCACTGATTCAGNTGGGCCATCAGATCCTACCTCCANGGGGT
TGGGAATTAGAAAACAGTGACCCTAGTGTAGTNTAGGCNACTTGAAGTGGAGGGCCCATACATTC
CGAGCCTTATCGGGCCATGTACACATGGAAGCAGGAAGANTGAAGGAGGGAGAGTAGAGGCCAG
AAACCCACCTGGGTTCTGTGTTCCCAATGNTAAGTCCCTGCCATGTTCCTGCTCTTCTGTGGTT
NGGATCTTCAAAGCTTCTCAAATNGGGGCAGTGGCCCTGGCAGCTTTTCAAATCCTYCCCAT
TTTATTGAAGCTGAAAGACCCCTTGACTAGAAC

SEQUENCE ID NO:19

PROGRAM blastn DATALIB nr Begin >1110.con

ATTGTTATTTTTCGTCACTACCTCCCCGGTCGGGAGTGGCTAATTTCGCGCGCCTGCTGCCTTCC
TTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCCGGAATCGAACCTGATTCCCCGTCACCCGT
GGTCACCATGGTTAGGCACGGCGACTACCATCGAAACTTAATAGCCGAGATCTCGAGAATTCTCG
AGATCTCCTTCMAATTATTACTTCANTTKCGGTAGTGATCAGNACNAGGCAGTTCTATTGATTC
TCTCCTTTCATTCTGAGTTTCTCCATAAATTAATTGGACCTAATCATGTTTENAATCCTGTCTTT
TAGGGGGNANTTGNACINTCAAGTGTTTAAACCGAGGGNCCGAGNATGATTTGGATTGGAGTGA
GAGCA

SEQUENCE ID NO: 20

PROGRAM blastn DATALIB nr Begin >1127.con

CAGANTTCTGGGTNAAAAGGACCTNANACATAATAAGTGGACTTNCATAAAACACTTACCAA
TGGANAAATGAACCCCTGGTCACCCCGATCTCACTAGTNCCTNCCCTGAAACCCGANATCTGA
GTGCTTTTCTCCTTTACTAACCTTNGTCCATCCTGCTCATGGGAATTAANGNTGTAAAATANG
CCTGGGGNACCTCGGRCCTCTNCCCTGGGNTCTGTGGGTGGGAGNACTGTGGGAAGCCGTWTCAT
CGCCCCCCTATGAGAGCCCTTTCTNCAGGGCCAGCCATGAACGTCCCCCATGTNATCAGNATCT
NCAGGCTACTGCTGTCTTCTTGGATWTTTAACTGCGGCGCGGGCCAGGCA CAGAAAARGGAGGT
GGCAAGATCCTTGAACAAAAGGAGCTATAAAAGGCGTTGGGGGAAGCAAGGCAAACGGCAGATT
AAACAGCAGGCCACCTCAAGGAACGTGACGC

SEQUENCE ID NO: 21

PROGRAM blastn DATABASE nr Begin >1147.con

CTCCACATCTGCCCCATCATTTACTTTTATNCCTTCNAGTNTNTAGNAGATAAAACATCCACGTC
GATCTNCTCTTAGACAAATCAANTACTTTAGGNATNTGATAGTCAGAGANTGGNTATCAAAATNGA
AAGGNATNTNGSTHGANCACTTAGTTTCYNCCTTNGNNGAGACCACTGGGNTCTNGASACCAGA
TTCTGGGGTNCIAATCTTANGGTAATCTTAGAGCCAAACACATGGGTCAINTTATSCCCCAAATT
AGCCACATCTEGTGGGCTATGGNGTCACCCCAAGAGCAGGAGGAGCATGONTGGATGGAAATCC
ATCTCCACCACTGGAAACCCCAATTTCTGAATGNATCACCTCTTAGAGTTTCTTGTYCATAAATA
GCAGGGAATTTAGGAATTTAGTTTCTTTTAAATAGTTTGGGCTTTTATCCACACTCTCAGCAGC
TTAGGATACTTTTCTCTTCAGCTCACTCTGAAACTCCCTCTGGA

SEQUENCE ID NO:22

PROGRAM blastn DATA13 nr Begin >1165.ccn

TCGAGATCTGTGGTAGTNACATGATATTCTGGCAGCTACTTTCATTATCACCTTATTTAAATAA
ATTTAAAGAAAAATGGCAGTATGTTTCTGTGRAGNCCAGCAGTACTCATTTTAAAGGACTCMAGA
GTTTCAGRNAAGTAAAAAGRAAGAGTAAATCATTTTCTAANTYTYWYTTCCAGAAATAACGA
TGTTGAGCATTAAAGTGGACTTCATTTCTACTCTTTCTMAGNTTATGTAGGCATAWAWATGTGTG
TGTATATACATAATATATGGGTACATCCTTAGAGAAGTTGGCTGGCTAGATAGACACACNTNAAAA
ATGGRATCATACTCTAATKCCATTENNANTTEANAAAATACATATTCAGANCCNCTGINTTATA
NACAGAGTAANTGAAA.

SEQUENCE ID NO: 23

PROGRAM blastn DATALIB nr Begin >119.csr

GACCCAGTAAAACCTTATCTCATGAGCATAAGGCTGAATGGGATTGACAGCCTACAGAACCCGGAT
TTTATCATGACGGCATTAGTGGGGGTGGGGCTTAGCTACTGAAAGTTTAAGGAGGTGAAAGGAA
AGCAACTTGTCCCTTACAGGGTCAAGCTAGGTCAAGGAAATTCCCAGGAGCGTGTGGAAGCTCTC
TACCTGATAGGTGAGCTCAAGCTTATGACCCCCCAAGCTTCTCCCCAAGCTTCCCTTCCACTGCT
TCCTCTTGATTGACTTCCACAGCAAGGTC

SEQUENCE ID NO: 24

PROGRAM blastn DATALIB nr Begin >128.con

CCATCAGGATTTACTGAGTAAAAATCTCAGGTNTTAACCATGCCCCCTAAAAATGTGCTATNCCAAA
GAGGAACAGGTTACTTGGGAGGAAAAAAGCTGCCTGGGNAACTCCCNCAAAATGTTTATTTTAAA
TAAAAATCGTNGATCGAAATATTTTNTAAAAGAACTTGGGGTNTAATATGGNATACTGCCCATCA
AACAAAAAAGGAAATAAACTTCNTTCCCATTTATAATAAGTTNCCCACCCCTTTACTATCAAGAT
TACAACTTATTGACCTTTTATGCTNGCTNGGTCTTTTGGGACTGCCTAATCCAATGTTTAAATT
TTCTANGTCTGNATTTCAATGTGGGTAGCAGTNATTTTCAA

SEQUENCE ID NO: 25

PROGRAM blastn DATALIB nr Begin >150.com

GAGTATCTGACAGGTAAGATTGCTTTTTAAAGTTGTTTTAAATGCATTACATGACTGAGAAAAGA
AAAATGCCACATTTTATTGTTGCAGTTTAAAAATTTCAATTTGNGTGAAACTAAACGTGAAACAAAA
GGGATAAATGTGTTTGNTTTTGTTTTGGTTTTACCTGTTGGGGTATTTTTTCTGAGTTTGTG
TAGAAACCCGTGTGGNTACACTGGGTAATCTTGTCAGGGNTACAAACCTGGGTCTTGANTTTGG
TTANTTGGNTTTANTTGGTGNACCCATGTACTTGCTCTTCCNTCCCAGAAACATAGCTTGGTAGG
CNAGGGTTAANCCAGTGTCCGCCANCCCATGTCCCTANCAAGCATCTTGTAAGTTTAATGCACA
ATCGTTCCNTCCCAGGATGGANTTATCATTATAAA

SEQUENCE IV) NC: 26

GAGAGGCGCAGGAGCCACAAATAAAGCAAGAGCCAGAAATCAGAAGNGGAGGAAGAAGAAAAGCAA
 GAAAAAGRAGRAAAYACGAGAAGAAACCCATGGRAGAGGAAGAGGANCAGANCMYAAAGCCTTGTCT
 GAAACCTACTCTGAGGCCCATCAGCTCTGCTCCATCTGTTTCCTCTGCCAGTGGNAATGCYACAC
 CTAACACTCTCTGGGGATGAGTCTCCCTGTGGTATTATTATTCTCATGAAAACCTACCAGATCAA
 CAGCAACCTGAGGAGCATAGGCCYAAAAATAGGACTAAGTCTTAAACTGGGTGCTTCCAATAGTCC
 TGCTCAGCCTAATTCTGTGAAGAGAAAAGAACTACCTGTAGATAGTGTCTTTAACAATTTGAGG
 ATCAAGACAGTGTATGACGTACCCCGAAAAAGGAAACTGTTCCCTTGGATTATGGTGAAGATGAT
 AAAAATNCAACCAAAGGCACTGTAAACACTGAAGAAAAGCGTAAACACATTAAAGAGTCTCATTGA
 GAAAATCCCTACAGCCAAACCTGAGCTCTTCGCTTATCCCTGGATTGGTCTATTGTGGATTCTA
 TACTGATGGAAACGTGJAATTAGACCATGGATTAAATAAGAAAATCATAGAATATATAGGTGAAGAA
 GAAGCTACATTAGTTGATTNGTTTGTTCAGGTTATGGCTCATAGTNCACCCAGAGCATT
 AGATGATGTTGCCATGGTACTTGTATGAAGAAGCAGAAGTTTTATAGTCAAAATGTGGAGATTAT
 TGATATATGAAACAGAAGCCAAAGAAAATTGGTCTTGTGAAGTAAAACTTTTATATTTAGAGTTC
 CATTTCAGATTCTTCTTGGCCACCTTTAAGGACTTGAATTTTCTTTGTCTTGAAGACAT
 TGTGAGATCTGTAAATTTTTTTTTTTGTAGAAAATGTGAATTTTTTGGTCCCTAATTTGTTGTT
 GCCCTGTGTACTCCCTGGTGTGTAAGTCACTGTGAATCCCTGGTTCCTTTTAACTCACCAGGTA
 CAAATTACTGGTATGTTTTATAAGCCGCGAGCTACTGTACACAGCCTATCTGATATAATCTTGTTC
 TGCTGATTGTCTTGTAAATAATTAALACSACTCCCAATTTTTCAGAAATTCACCTTAATA
 TTGAAATGTACTGTATAGGAACCAACATGAACAAATTTAATTGAAAAACCCAGTCCATAAACTATT
 ACCACCCCACTCTCTTTTCATCAGAAATGGCAAGCCCTTGTGAAGGCATGGAGTTTAAATTTGG
 AATGCAAAAATTAGCAGACAATCCATTCTACTGTATTTCTGTATGAATGTGTTTGTGAATGTAT
 GTGTAAAAGTCTTTCTTTCCCTAATTTGCTTGGTGGGCTCCTAAAAACATTTCCCACTAAAG
 AATAGAATTGTAAAGGAAAAGTGGTACTGTTCCCACTGAAATGTCTGTTATAATTAGGTTATTA
 GTTTCACAGAGCATGGTGTCTCTGTCTGTGAGCAATGTGGTTTGTCTAACTGGATGGGGTTTTCT
 TATTAATAAGATGGCTGCTTCAGCTTCTCTTTTAAAGGAATGTGGATCATAGTGATTTTCTCTT
 TAACTTTATTGCTCAGAAATGAGGCATATCCCTAAAAATCCTGGAGAGCTGTATTTAATGCATT
 TTGCACTAATTTGGTCTTACTTTAATTTCTATTGTATCTGTTTATTTAACAATAAATTCATCATAC
 CAAAAAGTGTAAAGTAAAAACCCCTTTTAAACAAAAAATAAATAAATTAAGGCAATTT
 GACAGACAGTGAGAGTTTACAAACATGATAGGTATTCTGCTCGGCAATTTGTAAGTTTACATGT
 TATTTAAGGATAAAGGTAAATCATTCAGGCGAGTTACCAACCATAACTATTGTTTTTCATTTT
 GTCTGTAGAAAGCTTTATATCTTGTTTACCTTGGCTCATTAGTGTTTAAAAATGTACTGATGAT
 GTGCTTAGAGAAATTCCTGGGGCTTTCTTCTGTGTAGATCAGAAATTTCAACAGGGAGTAAATTA
 CCTGAARACGTAAGAAGTTTTTAAACAGCTTTTCAACAAATTAGATGCAACTGTTCCCATGTCTG
 AGGTACTTATTTAAAGAAAGGTAAAGATTGGCCTGTAGAAAAAGCATAATGTGAGCTTTGGAT
 TACTGGATTTTTTTTTTTTTTAAACACACCTGGAGAGGACATTTGAAAAACATGTTCTTACCTTC
 GAACCCCTGATGTGGTTCCATTATGTAAATATTTCAAAATATTAATAATGTATATATTTGAAAAAAA
 AAAAAAAAAAAAAATTCCTGCGGCCGCAAGGCAATTC

SEQUENCE ID No: 27

PROGRAM blastn DATALIB nr Begin >170.con

ATTGGAGCTCCACCGCGGTGGCGGGCGCTCTAGNAACTAGTGGATCCCCCGGGCTGCAGGAATTC
TCCAGATCTCCCCCAAGTAAATGAATGAAAAAAGAACACCAACATACAGATGATATAATAAGC
CAGCCATGGATGACCTTATAGCACCCCTGTATTTATACAGAACCACCAAGGAGGATAGTCATGACAA
CATTGACACTGATCATGATNCCAGCATTGAGAATTGAGTNCAGGGCTCTCTGGCCACAGTCTCG
GTATCTTCTGTGNATGGGGTATAGATTAACCTGTCCATCCTTCCGGGNATAAANCTGACTGACTT
AATGGTANCCACGACCAACCAACCATKAGAGAGTCAAGGGACTAAAGAGCATGATCAACATGCT
TGGCNCCATATTTCAATNTCANCTCCTCATCTTCTTCTCATCTTCTTCTCCACCACCTNCCGGGAG
TTAACCCCTGGGGTCGTCCATTAGATAATGGCTCA

SEQUENCE # NO: 28

ACGGTGGCTTCAGTGTGGCTGACACAGCAGCATGGTCTTGACAAGTTTTCTTCATCTACCAAAA
ATCCCACTTGGTAATAGAGACTTTACTCTACCTATCAAAACCAAAAATGTCCCAATTAGGGGGG
GACATGTTGTACATGTTAGGATCAATCAATAACCAAGATTATAAGGTGAGGAAGATGCCCTA
ACTGATTCTTTTGTCTCTCACTCTTGTTGGTTCCAGGGACCAAGTGGGGTCAATCTTCTGGTSSG
CCTCTCCAGGTCTCTTCCAGGCGGGTCAAGACCTACTCCCTCTGAGGCGSACCATGGTTAGAA
GAGGTGTCTAAGAGCGTCCSGGTCAAGCAGCCCAAGGCTCAGCTGCTCTCTGAGCCAGTTCT
CCAGGCTCTCCACCACTGCCATCTCCAGCCAGCATCACTTTCCAAAGGAATGCAATTCCTCA
CTCTCAGCCTGTGCGAGTGGGTGTGGTCCAGCCCTGCAACCAAGCCTTTGTCCCTGCCAGTCC
TATCCTGTGGCCAACTGGAAATGCCCTATCCAGCCCTTAATGTGGCTGTGGTGGGCATCACTCCCTC
CCAGATGGTGGCCAACTGTTTGGCACTGCAGGCCAACCTCAGGCTGCCCATCCCATCAGTCAC
CCAGCCTGGTCAGGCAGCAGACATTCCTCACTACAGGCAAGCAGTGGTACCAACAGTCCCTTC
TTTAAGCCTCTCTCAGCACCTCAACGGTTCTCCAGCTTCAATGGTGTAGATGATGGCAGGT
GGCCTCAGCAGACAGGCATACAGAGGTTCTACAGGCCCTGCTGCTGAGTGGATCTTTTGAAGCCC
AGTGGGCTGCCATTAGAAATAAGTCCAAGCAGCGTACTAATCCCTCCCTTACCAACCCCTTCTCC
AGTGACTTACAGAAACCTTTGAAATTGAATTTAAGCAATCAATTATGGCTATGTATCTTGTCCA
TACCAGACAGGGAGCAGGGGGTAGCGGTCAAGGAGCAAAACAGAYTTTGTCTCTGATTAGTAC
TCTTTTCACTAATCCCAAGGTCCCAAGGAACAAGTCCAGGCCCAAGTACTGTGAGGGCTGATT
TTCAAGACATGGGAAAAGCAATTCCTAGAGAAAAGCTGCTTGCATTAGGCTAAAGAAAGTCAA
GGAAATGTTGCTTCTGTACTCCCTCTTCCCTTACCCCTTACAAATCTCTGGCAACAGAGAGGC
AAAGTATCTGAACAAGATCTATATCCCAAGCAGATTTACTGAAATGTAAACACACAGGAAGC
AAAGCAATGTCCCTTGTCTTTCAGGCCATTCACTGCTCTCTGTCACTAGTGGCTGTATTAGA
GATCAAGAAAGAGTGGTTTGTGCTCAGGCTGGGAACAGAGAGGCACGCTATGCTGCCAGAATTCCT
AGGAGGCCATATCAGCACTGCCCAAGCAGAGCTATATTTGGGGGAGAGGTTGAGCTTCCATTTT
GAGTAAACAGAAATAATATATATATCAAAAGCCAAATCTTTATTTTATGCAATTTAGAATAT
TTTAAATAGTTCTCAGATTTAAGAAAGTTGTATGAGTTGTAGTAATCTTGCCTAAGGTAAGGG
GCTAGTTGTAAAGAAATGTACATRAGATTGATTTATCATTGATGCCCTACTGAAATAAAAAGAGGA
AAGGCTGGAAGCATGCGACAGGATCCCTAGCTTGTCTTCTGTCACTCATTGATTTTAAATAAAG
CATTGCCAACAACTATGCTTATGACCAATACAGTCACIAGGTTGTAGTTTTTTTTTAAATAAAG
GAAAAGCAGTATTGTCTGGTTTTTAAACCTATGATGGAATTTAATGTCAATTATTTAATGGAAT
CAATCCGAAATATGCTCTATAGAGAATATATCTTTATATATGCTGCACTTCTTATGTTAATC
CTTAAACACTAAGGTAACATGACATAATCATACCATAGAAGGGAACACAGGTTACCATATTGGT
TGTAATATGGGTCTTGGTGGGTCTTGTCTTATCCCTTAAATTTTGTTCCTATGAGTTTTGTGGG
ATGGGGATTCTGGTTTTTATTAGCTTGTGTGTGTCTCTCTCCCCCAACCCCCCTTTTGGTGAGAA
CATCCCCCTGACAGTTGCGCTCTTGACCTCGGATAACAATAAGAGAGTCACTCTCATTTTTAC
TTTTGAACGTTGGCGCTTACAATCAATGTAAGTTATATATATTTGTACTGATGAAAATTTATAA
TCTGCTTTAACAAAAATAATGTTTCATGGTAG

89. cm

SEQUENCE ID NO: 29

GGCAGCTATTTACATGGCCTCACAGGCATCAGCTGAAAAGAGGACCCMAAAAGAAATGGAGATA
TTGCTGGTGTGGCTGATGTTACAATCAGRCAGTTCTATAGACTGATCTATCCTCGAGCCCAGAT
CTGTTCCCTTACAGACTTCMAATTXGACACCCCACTGGACAAACTACCAAGCTATAAATTGAGGC
AGYTAACGTGMAAATCTTGANTACMAAACTTNCCTGTTGTACATAGCCTATACMAAATGCTGGG
TTCAGCCTTTTATAAGGNAAAACTAAGACATGGNTACGCATTCCAGGGCTKGANTACTTATTC
TTGGCATCTCTGTATGTA

SEJWENK 2) No. 30

PROGRAM blastn DATALIB nr Begin >237.con

AAAGGGCTAACCCAGCCACTGCACCAAAATTAGTCCTTACATTATAATACTCTGGCCATTGGAAGA
GAAAAATGGGAAAATTCAACAATTGAAAGACTATGATCCCTCTGGCTCATCATCTACTGACCAG
AATGAAGTCCTGAAGGATTTCTTCTGTTATGTTATCTACCCAGCTAATCTCAACAAGAGGAGC
TGGAAAGAACAAAGCCCCATGAAGCTACCCCTAGACCCAGAAAGCCAAAGAACAGGGCCAAAGAAA
TGAACAGCAGACAAAGCCTGAAATAGAAAGTGGNACAGACATGTGGNAAGACCAAGTACACCCAGTT
NGGTGGTAAAGATTCCGATATCAAGCTTATCGATACCG

Sequence ID No: 31

PROGRAM blastn DATALIB nr Begin >265-325.con

AGTACATGGTTTCTGNCACCCCSACCCCTTTCCCCATCTCTACCGGTGATAGTCTCTCAGN
TAGTAGACCTTTTCTNGTTTAGRCAGGGCCACNTTTTAAAAAAGTCCAGACGGGTACCCCTCCATG
TKGMAGGCGACGTGGGCGCTGGATCACTCAACTGANTGTCATNKGANTGGTGCCCCCAGAGTGAGG
ACAATGGTGNAGCCCTCCTAAGGCCCTNCCCTGAGTGTCCCTCCTTCATGAAGATGATTCTGAGGH
TTCCCGAGGCGCTNCACCCCTTCTTKGAAARCCCATAGNAGTTTCATATGNACTNCTCTNCTATGCTCA
CCAAACTCTNCCTTCATCATACTGGGGGATGTGTGT

SEQUENCE ID No: 32

PROGRAM blastn DATALIB nr Begin >293.con

GTGCATGTAATTACAGTTACGATATATGAAACGTACAAAATATTATGAGTATATAATATGGGGAG
ACTTAATCTAGTTTGGGGGATCAGGGCACATTTCTCTAAGAAAGTGACATTTGAATTGAGCTCTG
AAGGATAAATAGACATTACCCAGAAGAATAAAATGATGGGGAAGAAGGAGGACATTTTCCGTAGA
TTTCCAGTGGCCCCNCTTGATCCCTTATCCACTCATCACTNAGGAGGATATTAATKCTATAGAA
ATGGRAGRAAGACCTAAAGAGACCCCTNATATCTCGAGAGGATCCAGCMAAATTCCAGAGACACA
ACAWTAAGAAACTNGGAAGGAACAGAAAAGGCTTTNNAGGNAAAAGCAAAGACAAGGAAATTNNNN
NAGNACGGAGAGAAAAGAGAGAGGGGAGCCTTNAAGGGNACGAGAAAGGCGAGNACGGGGACGAGAA
AGGGNAAGAGNACGTAACG

SEQUENCE ID NO. 33

295 245 C-TT

CGAAATAAATGAGATCTCAGTGGTGGTATGGATTGGACTGATCTCTGTAACTGTGTNTGGAAAA
GGACCCGAAATGAAAGCCAGATCCCAGTAAGGGGTAGAGAGGGGCCAAGAGAACTGAACATCTG
GGCTGCCCCGAGAAATCAAAGTCTAGGAAGTAAGAGGTAAAGAGTGTACTACAGGGGACATACCCCA
ATCTCTTGGTTCCCTCCCTCTNCCTTCCTCTCCAGAGACCCAGGTCCCTGGGACTATNTTGGAT
CTGTCTCTGAAGCTGAAAAACAAAAGGCAGAGGAGACAGTCCGNTCTAAGTGACCAATCTCAAGC
CAGCTTGGTCAGAANTCCTAA

SEQUENCE ID No. 34

PROGRAM blastn DATALIB nr Begin >295-43.com

AAATCCAGTGCAGGCCAACATTATGTGGAATAGAAACAGGGCTCCTGCTAGGAGATTGANATTCT
GGCTTTCCTTTGGAAACCCCTCACTGACTCATCGCCCTGAANCAGGANCCANACAGGTNCCAAGGC
TCCCTGCTCCTNTCCCTNCCCAGGGCGAGATAGGAARCCGGAARCCTGGGCAGGCTGARCCCA
NCCGACTGGAACCAGGGNAGAN CCTGTGGGTGGGTGGNAGGGAGGGAAGGAGGCCAGATTCTTC
AGAACTGGGGRAGAGAACAGGTTTTGGAAGTTCCGGCAGGGTTTTGGGTTTCACACTCATGGTTTC
ATCANVCCCTGGAGGGTTNCACACTCCTGGTRCANTTTTGTANTCGTNCTTTGAANACARNCCG
CTTCCTTTC AACCTCCNCTAAAAAGTTTTGATNTTTTAAGG

295 48 cm

SEQUENCE IN: 35

ACCAAGAGCCCCCAGTTTATGNTAACTCTCATGACAAACACAAATTTTAGTACCTCTCACTACCAA
CTATCCAGGAACCCAGGANTCACCTATTACTACGGTTCCAGCAGAATGGGAATCCCATTCTCGGAT
ATCCAGGGTAAATCCCTGACCATGTGAGAGGAATCCTAGTGGCCCAACAACTCAGCCCCCTGACT
CCTCCTCAANGGCTCTGCCAAGTCAACAAAAAATCCTCTACATTTACACTATCTGTAAAGCCAA
AGACCAGCCTCAACCTAAATCTCCATCAATAAGGGAATGCTTGGATAAGTAAAAATTATGCAGCT
CTAGGAAGGAATGAAGAATGTCTAT

SEQUENCE EV. No: 36

PROGRAM blastn DATALIB nr Begin >310.con

AAAGGGAACAAAAGCTGGTACCGGGCCCCCCCCCTCGAGGTCGACGGTATCGATAAGCTGGATATCG
AATCCTCGAGATCTACCTAAAAAATAAATTAACCTCCCAATGTGGGAGTCTACTCTGTTCCC
TCCTNGTNTTTATTNCTGNTTACTTTTCTAATATGGTTAAAAATGTGTAANCAATATGTGTCTTT
NACTNKGKGTGAACAATTTTCTATTATAAATCTWAGAAAAATATNCTATGGNTATGAGATAT
TKGATTCGAAGTGCTTGTAAATTTACTYCTCAATGTCCCTGATGTGKANATTXGTTNCTAGTG
TTYCACTATTTAAAAAACAGNAATATCTGTCTNTATGCTNAGAGCTTNTYCAGTTTTCAAATTA
TTNCTTAGGGTAAAAATCCTAGAAGTAGAATTTTGGGGCAAATTAATCTACATATTTATAATTGT
CTTCGTATTCCAAATCTCGTTTCCAAAAGCTTATATCAATTTGTACTTAACACCAG

SEQUENCE ID NO: 37

PROGRAM blastn DATALIB nr Begin >122.csn

ATTTAAGATGACTGGGGGTCTCTNCCTAATCCCATACTCCACTGGAGAGGANAAAGTCCCAAAGCT
TGGTCTAGTTARGGTNGNTGGGGACCCCTCCCAAGAGCTGNAGAAGCAGAGATAAGNAGAGCCTNC
TNTAAATCCACATGGNCCTYCCAAGGNTCTCATCTCTAGGACCTACCCTNCTCAGTCTACTT
ACTTGTCTYCTGANATGCTTCTNGAGGGGNAGAAAACAAAGGAAGAGTAATAACAAGCAGNAGA
AACTGCACAGAATGNAAAATAAGTCCATAGGAGAATGTTGNAAATAGAATCATCCNCCTTTACAT
ATTGTCACTCCAGGAAAAGTGCCTAAGAACCCTCATTCCTCTAGATACAMTTCCTGTAGGATCTY
CCCAGACTTCCTCCCTTAAGCACGTCAGTATTCTCCTTATTCTCCCTTCATTTCAACCT

SEQUENCE ID NO 38

PROGRAM blastn DATALIB nr Begin >129-1127.con

CSAGATCTGCCCCAGCCCCACATTTCTTTTGTGAATGAGTAGAGAAGACTGAGAAGTATCACTCA
CCCCGTGATGTGGTTTGTCCCTTTTCCAGCCAGTGTGTTGGTAATAAAAAGTCACCTTTTCAGAGCTT
TGGTCCCCCGTAATGCCCGTCTTTCCTGTGTCCAGGAATAACCTTTGNTACTAGGCAGTCCTCTGA
AAGATTTGTAGAAGGTTAAAGTGGAAGGGGACTTGGAAAGCTCATAGAATCCATGCCCTCTTCTTTT
AGCATCAAGGAATTAGAAGTCCTGAGAGATGAAGAATGTTGTCTTCCCAACTCAAACCCATTCT
TGAAGCCATTTCCCTCGTTACTGNATTGGCCACAACCCCTTCCCCCTTGNTATCCTCATCCTGCTA
ATGCTGTTTTTAATGGCCCTGCCAGTCTGGATTGTCTTTGGCAACCAACAATTTTGCTTCACAA
GATTCCTACTTAAGCGAAGAGAGGGGCTCCTCATTTNTCACTTGTACAAGAGCAGGGGCTGGTCAG
CTTTACACAGGTGTCAATGAACCCCTCACAANCCAGANTTNCATGTTGGCCCTCAGGAGGGCTTCN
AGGTCCAACATCTCGACGTAAGGAGCGTTCCCAAGTTCTTTCATGCTCAGATAACAGTNCCTAACTN
CAGCTGTTTCATCCCAATCCCTANTTGAGGTCTTAACATCTATTCCATTTTXCCNACAGGGGTT
ATNCTGTTAACCCCTCTNCAACAGANTTAGANCTGACTGATNCACTTCCTAG

SEQUENCE ID NO:39

PROGRAM blastn DATALIB nr Begin >424-216.con

TCATACCTTGTATAGTTCKNTAAGATAATCACTCTCTCACTCAGACATNNGGNGRARNGCCCNTCG
ATCACTTGGGANAGGNGACTTGCMATGTTTAATGATTGTCAACCGNANAANTAAGCTNACAGGGC
AAAAACAGCCTYANGTCAGTTCTNTCTCCCTAATCCTCTAGRAKNAATCMNAWENTRNACTCT
GNTCTGTGCCATNANANATNTTNCANTGTATTTATGNACTCCACATNGAGTACACCTCACTAA
WTNTNCTNCTGGGNAACNCCCS CMCCANTTTTTTNTTGTGNTGANANACARCAATGCTGGCATACTG
TG

SEQUENCE IN NO: 40

PROGRAM blastn DATALIB nr Begin >424-237.con

CCAGACTTTTCATAACTNGTGTTATTATGAAGATTAGAGTNGCTGAAGCTTACTGGATTAGAAGAGN
ACGAGGGGGTAGCTGCCCAATATATTCTAATTTCTCTKGAGGACCAACCAATNGGMAGAGTGTG
TCTGATAGGGGAAAAGGAAGAGTTGGAAGGNATCTTAGCCTCTAGGANAAAAGAACCATTTTTATT
GGCCACCAAAGTTACATCTAGTXGCCCTACAAATTTATNTCCAAACTCCTTATCCTGCCAATTCAG
GGTCCTGNAAACTGATGCCAAACTATAGTTTAGTCTNCTATCACATGACTGCATTATACATACCC
AATTATCTGGGMAAACAGACCTGATCCAAACACAGTTKGGTNGCTTTCTTNCCTTNCCTTKGTTT
AGCCTGTGCCCTCTACTNGGGGTGCTCTTKGATTGCTCCAG

SEQUENCE ID NO: 41

PROGRAM blastn DATALIB nr Begin >450.con

TTTTTTTCCACCAGACTTACCAAATTTTAGATGNATGGAAGAACTGTAAATNCCATAAAGNTAA
TCTATNCATNGACCCCCACCAATTATGATAGAGATCATNTGGTGANTAAATGAAAGATGAAACTCTC
AGCTGGGAAAGTAANAAGGAATAGGATGTAAGTATGAGCTCCTGTTTTTTATTATNTTTATGGAT
GCCCCCTCAGAAAAATATGNAANGGGTAACTGACTNGGAAATGGGTNTTTTATGNATAGTAAGT
CCCACTCAGGAGGTTT

SEQUENCE ID NO. 42

PROGRAM blastn DATALIB nr Begin >462.con

TCCAGATCTAAAGCAGATGNAGACTTTNCACNAAATAAAATTIACGCTTTTTTCTGTGANATAA
GTTNCGAGAAGCAAAGCTTTKGATTNCTRNACGAGTYCAGTGGATTATYCTNAGNACTAGAGTKG
NKGTKGAAGNCATGGNACATTTATATAGTYYTTTCAGTTCTACACTAAATGATGGAAGAATGAGA
AATCCTATATGACAAATAGAAAAGTYCATYCTYCATAATTGAGAACATTGAGCAGTTGGATTACC
AAGATCTCGA

SEQUENCE ID NO: 43

PROGRAM blastn DATALIB nr Begin >485.con

CTTAGTTTTAGACTAGTTTCATTATACTACCAGTTTCTAATATGTTGGTTTTTATTCACTATTT
GATATATTTGTTTTAATATATGTTCTTGTTTTAGCAGGTAAAAGAATCATAACAAATGTTTTTAA
AAGAACATTATTTCTTTAATAACTGTCTTTTATGCATTGCGCATGCCAATTTTTTCATTAA
CATCTTGGGTATTTTATAAAAAGAGGGAAAGCTCAATGTTTAAACAGGTAGCTTTTCTTAGGAGCT
AAATTAAATATTTAACAATCTCCTTCCTTCNCCCTTCCCCATCCCTCAAAGNATGGGTGNANT
TATCTTTAACTTTTGGGCTNGCATCCNTGNAAGCTTATGGNTANTCATAGTCTNACMAAACTAGG
GTCAACCAACTTGGCAGCAGAAATAATCTAGTCTTACTGTGATAACTACCCAATTACTTTATTAT
TTTTCCAGTTTCAGTTCCAAATGTTTTGTGGNAANAATTTTNCCTGTTGTGATTTTCCAAGCTT
AGAGGGGGAAACCAACTTCCAGTGTGGAGAGCACTGNATAGTTTATGNATTGTGTAAA

1553 SEQUENCE ID NO. 44

TTTTTTTAAACGAAAAAATTACTGACACGACCTTTCTAAGTGTATTTATTTAC
TAAATGATCTTTAAATTTTAACTTACCTACCTAGGAGTTGATTAATTTTCTGCT
CAACAGCTAGTAACTATGATAGGCTGGGCTCAGTGGCTAAGCTTTAACTCCGCA
CTCTGGGCTTACGCTGGGCTCTCTCTTCTGAGTCTAGGATTTGAGCTAGCTCTGCTA
ACATGGAACCTCTCTTACTAAATTAAGAAATAGCTGGGCTGGTCTGCTGGCT
ACTTGTAGTCTAGTACTCTGGAAGCTTGAGGCTAGGAGGATCTCT

1560 SEQUENCE ID NO. 45

GTTCTCTAGCTTCACTGGGAGGCTGTCTGGGAGAAATGCTCTTCACTAGTAAAGAA
TCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGAGGCTCTCTCACTAGT
GAGATCTCTCTCTCACTAGGATAGGAGGCTCTCTCTCACTAGGATAGGATCTCTCTCT
CTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCT
TTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCT
ACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGAT
GCTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGAT
CTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCT

1562 SEQUENCE ID NO. 46

CAACCTCTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGAT
ACCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
GAGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCT
ACCTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGAT
CACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGAT
NTCTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGAT
GGTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCTCT

1570 SEQUENCE ID NO. 47

AGCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
GGGCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
CTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
TAACTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
TAACTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
GAGCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
CACTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
CACTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
ANTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
TCCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCTCT

1571 SEQUENCE ID NO. 48

AAAACTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
AGGCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
TCTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
GGTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
GTTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
CACTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCACT
CTCTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCTCT
AGTCTCTCTCTCTCTCACTAGGATAGGATCTCTCTCTCACTAGGATAGGATCTCTCTCTCTCT

ATCAAGCGACAGTGTGTATTCCTATACGCGATTTTATATGTGTATATAAAAAATTA
 TTTGTTACACATGTTGATCTATTAAGTGTGTGTGTGTATAGATGATTTCTCTCCCTT
 GAAGGTGAAGCAAGCGTCCTCTTATTAAGCGATAAACTTGGGTCTATATGATACCTTC
 GGAATAAATGATTTATCTCCACTTGGAAATCTCAAAAACAGTACATATATATTTCT
 TTTCTTTTATGTTTACGGGTCTTCTCTGTGTGCGCCAGGCTGGAGTGTGAGTAGTGTGA
 CTAAGTCAACAGNCTTAACTCCAGGTCAGNATATCTTCTGCCCGAGNCTCTGA
 GAAATGGGAC

CAAAAATCAAAGGGAAGTCTGAACCCCTGCGCAACCTCTCCATTCGCCATCTGCGCTGCTG
 GTGNGTCTCTCTCTCCTCACAGTACCTCTCGAAAAGTTCCAGAAATTCAGTTAATACAGAAATA
 TGGGTGAATTTTCAACCTCTAGTTTAAAGATGAAGAGTCTCCGTTGGTTTAAACCACTTC
 AGCTAACCTCTCTGGTAACCTGTAGCTCTCGACAGCTCTCCAGCTCTCAATGAAAATCTCTCTGT
 GACCACGCTCTAAGCTCTGTATGGGGACAGAAATCTCTGGGCTATCAATCTCTCTCTGA
 NCTCGGCCCTCAAACCTGGTAGTTTCTGCACCGAGGGACACAGTCCACTCTGCGATGAAGTA
 TGTTCAAAAATCGTTCTTTAGGGAACCTCTCGAAAGTCCAAATAGTGNAAAGGTGGTCAA
 GGAAGGATTTGGAGGAAGCTGTAAGTCAAGNCGGGAATCTGATTTGGNTAGNTGTGG
 ATGTAAGGAAATCACTGCGGCT

GGAAAGAGGGTCTCTTAACTCCCTGCTGCTGTAATAATCCGCTTTTCTCTCTTTCTGGTGG
GAGACAGAGTCTCTGCTCTGTAGCTCAGGCTGGAGTGGCTGAGGCAC

AGTCCGAGCTAAGCTCAGGAGGCTGGGGCAGGAAGATAGCTTGAGCTGGGAGTTAGAGGCT
GTGTGAGCTATATATCACTCACTGCACTCCAGCTGGGCAACACAGCGAGACCTTAAAC
TAAAAAAGCAAAAGAAAAAATATATGTATCTATTTTGGAAATTTCAAAGTGGGAGATA
AATCATTTTTCAGCAGTATCTTGAAACCCAAAGTTATGGCTAAATAAAGGTGTGCTT
TCTTCACCTTCAAAGCGGGAGAGAAATCATCATACACACACACACTTATACATACAC
ATATATACAAAAATACATTTTAAATACACATATAAAACATGGAGTATAGGCATATACAC
CTGTGCTTGAATAAATACAGGATCC

TATATTTNACCAAGCAACAGTGTGTTATGCCCTATACTCCATGTTTATATGTGTGTATTTAA
AAAATGTA?T?TGTTATATATGTGTACCATTAAGCTGTGTGTGTGTGTACGATGATTTCTCT
CCCGGTNTTGAAGGTGAAGAAGCAACACCTTTATTTAAGCACTAAACTTTGGGTTTCT?AGA
TACTGTCTGGAAAAATGATTATCTCCCACTTTGAAATTCCAAATACGTATCATATATTT
TATTTCTTTTCTTTT?AGT?NAGGGCTTGTGTGTGTGCCCAGGCTGGAGTGCAGT
AGTGTGATCATAGTTCACACAGGCTCTAACTCCAGGCTCAAGCTATCTTCTCTGCCCGAG
NCTCTTGAGTAGGTGGGAGT

3537 SEQUENCE ID NO. 54

CTGCAGTAAGCCACCTTCACTGCTTACTCTAGCTGGATGACAGAGAGAGATCTCTT
CTTTGGAAGAAAAAAGAAAAAGAAAAAGAAATATGGCTATGCTCTTATATATAG
AAGGGGTCACTATTAATCTGAAAAATGGATCTCTTGTGGGCTTCAACAGGCAACA
GCCACAGGTACCTACCTGAAGCTTCCCTGGGT TTAATGGCTGG? AGTATGTTCTAAT?
GTTCAAG? ACCCATGTCAC? ACTGGTGGTACAGAAATGGAAATCTCACTCTCTT? AAAT
CTCTTTTATTTTAAAA? GAATTAATCTA? TACATTACCTTATAAAAAGTAGGTAACCTA
A.TTTTGGTTTAAAAAGTGAATGAGGGCAGATGCAAGTGG? TCACACCTA.TTAATCTCTA
AATACCTTGGAGAGGGCAAGGTAGGAGGATTCCTTGGAGCCAGGAGTCCAAAGACCAAG
CTAGGGGAATATTATAGAAGTCTCTCTACAA? AAA? AAT

3538 SEQUENCE ID NO. 55

CTGCA?GAAGCTTTT?TTT?CTTT?GGNGGAGACAGAGTCTTCTCTGTCTCA? CCCAGGCT
GGGGTGCAGTGGACAGTCTATAGCTCACTGCAACCTTGAATCTCTTGGNTCATGCCATCC
TCCCACCTTCAAGCTCTCAAGTAGCTAGAACTACAGGTGTGCAACCACTGCTGACTAAC
CTGT.TAATTNNGGGGAGAGAGACCT? TCTTGGTATATGGCTAGGCTGGTCTTGAAGCTC
TGGGNT? CAAGCAATCTCTTACCTTGGCTCT? CAAGGTA? TGGGAAT? ATAGGTCT
GAGCTACTTGCATCTGGCTCAATCTCACTTTAAAAAT? CAAATTAGGTTACCTTACTTT
TAAAGGCTAT? GTATTCAATTAATCTTTAAAAAATAAAACCGATTTGGGAAAGNGTGA
GA? TCACTTCTGTAAACCACTGCTGTAAATGGGTCCCGAAACAGGTAGAACATACTC
CCAGGCA.TAACCCCAAGGGAG? GT? CAAGTCCCT? C

3539 SEQUENCE ID NO. 56

GGATCTGGTTCTTAAAAACAGAAAAA.TTACTGATAGTACATTTTCTAAGTGTATTA
TTGTATTAATGGATCTTAAATTAATCTTCACTACTGACATAGGATTTGAGTAACTTG
TGTGGTCAAAATAGCTGTAGTGAAGTACAGGCTGGGCGAGTGGATCAAGCTTGAAT?
CCAGGCTCTTGGAGGCTGAGGCAAGCAGATCACTTGAGGTGAGGAGTTTGAGACCAAGCC
TGGCTAACATGGAATACTCTCTCTTAAATAAATAAATAAATAGCTGGCGCTGGTGG
GTGGGCTCTTGTAGTCCCAAGCTACTTGGAGGGTGGAGGCAGGAGGAATCGCTTGGTCTC
CCGGAGGGAGAGGTT? TGTGTAGCTGAGATCACGCCACT? GCACTCCAGGCTGGGNA
CAAAAGGCAACCTTTCTCAAAAAAATTAATAAATAAAGTGAATAGTATAGGATTTGGA
CCCTAGACATCTTTTCTCAAGACC

3530 SEQUENCE ID NO. 57

CTGCAGTCTCAACCTTGTCTGGGATCAACAACTCTCCACCTCAGGCTTCAAAGTA
GATAGAACTACGGCATGCACTACCATGCCATAATTTTAAAAAATAATTTTTCAGA
GATGAGATCTCACTGTGTTCCGAGGTTGTCCGAACTCCTGGACTCAAGGATCTCTCC
CAGCTTGGGCTGCCAAAGTGTGGGATTACAGGCATGAGCCACTATGCTTGGCCAACAC
TT.TTTT.TTTT.TTAAACAAGACGGAGTCTNGTTCTGTGCGCCAGACTGGAGTGCAGGG
GCCT/NATCTTGGCT.CACTTGAAGCTTCCCTCCAGGGTTCATGCCGTTCTCTGNC?
CAGGCTCCCAAGTNGGTGGGACTACAGGATCTGCACCAACCCGGTTATTNTTGGGTT
TGNNGNAGGGAAGGGTTTACCATGTTAGGCAGGATGACTTCGGACTTCCNGACCCAG
ATCACCTGCTGGCTCCCA

3592 SEQUENCE ID NO. 58

GAATTCGACGAGCTAGCTGGGCAACACAGTGGAGCTGATCTACAAAAAATTTAAA
TTAGCTAAAGTTGATGGACATGCTTTCAGTCCGAGCTACTCAGGAGGCTGGGCAAGAA
GATAGCTTGAAGCTTGGGAGTTAGAGGCTGTGTAGCTATGATCACTACTGCACTGAG
CTTGGGCAACACAGCTAGACCTTAAACTAAAAAGAAAAGAAAAAATAATATGTAC
GTATTTGGGAAATTTCAAAGTGGGAGATAAATCAATTTTTCAGACAGTNTCTTAAACCC
AAAGTTATGCTTAAATAAAGCTGTGCTTTCTTCACTTCAAANGCGGAGAGGATCA
TCAATTCACACACACACATTCATCAATTTTCAAAATYCAATTTTNTAATACACAC
CAATTTAACTGGGTTTTG

3593 SEQUENCE ID NO. 59

GGATCTGTCTCTTAAACGAAAAAATTTACTGATAGTACATTTGTTCTAAGTGTATTA
TTGTAATAATGGATCAATTAATTAATCTTCAATAGTACATAGGAGTTGAGTAACTTG
TGTGCTCAAACTAGCTAGTAACTGATGAGTAGGCTGGGCTCAGTGGCTCAAGCTGTAACT
CCAGCACTCTGGGAGCTGAGGCAAGGCAATCACTTCACTTCAAGGATTTGAGACCAAGCT
TGGTCAACATGGTAAACCTCTCTCTACTTAAATAACAAAACTAGCTGGGCTGTGCTG
TTCTGCACTGTAGTCCGAGCTACTGGGAAGCTTGAGGCAAGGAGCAATCTCTGATCCC
TGGGAGGCAAGCTTGTGCTGCA?GCTGAGATCACTTCTCTGACTCAGCTGGGAA
CAAA?GATGAGATCTTCTCAAAAAAATAAAA?TAAANGTGAAGTAGGATTTGGA
CCCCAGACATCTTCTCTGAGGACCTGGATTC

3594 SEQUENCE ID NO. 60

GAATTCCTGGCTGAGTATCTCTCTCAGCTTCTGCAATTTGGGATTTGAGTGT
TGAGGCACTGTGCTAGGCTGATATATCTATTTTAACTGCTGTAATCTCATTGTAT
GAAATTTACTCTCTAGCTTAAATTTGATAGCACTTTAAATTTTCTAATTTCTAG
ATCTTTAACTTAAATTTCTGAAAGTATAGTATGGCAATTTAGGTATGATATGATCTT
NTCTCTCTCTGACTCTTTTATAGTTATGGCTGTGCACTGCTTCTCTTCTATCT
AATGATACAGAGCTTCTTAAAGAAAAAGTTCAAGCTTGGGAAAAAAGTAAATGT
CAACTTTCAGGGAATAAAGTAAATTTG

3595 SEQUENCE ID NO. 61

TCAAGTACCTCCCTGAATGGACTGCTGGCTCATCTTGGCTGTGATTCAGTATATGCTA
AAACCCAGACTGATAAATTTGTTGTCACAGGAATGCCCACTGGAGTGTCTTCTTCT
CATCTCTTATCTTGAATTAGCAAAATGGTAACGTGTACATCCCATAACTCTTCACTAA
ATCTTTAAATTAGCTATAGTAACTTTCTATTTGAAGATTTGGCTGGGCACTGCTAGCTCA
TGCTTGTATCTTAGCACTTGGGAGGCTGAGGCGGGCAGATCACTAAGCCCCAGAGTTT
AAGACCAAGCTTGGGCAACATGGCAAAACCTCGTATCTACGAAAAATACAAAAATTTG?CG
GGTATG

[illegible]

TAAACACAGGGCTCATGAGGGCACTAATCA.AATGAGATATGCGTGCTGGAGTCSAAAGTG
 GACCTTCCAGTAAATGGAAATCATTCACACCCACCAAAATTCGAGATCAGAGAGTGAAA
 CAGTAATGTAGTCCACAGCAACGTTACAGGTTTAAACCTTCCCTGAAAAAAAATAC
 CAGATTTAAAGAATTAATAATTTCCACCAAAACHTATTAGAAATATGTGATGCT
 TCCCAACATTAGATAATACATTCACCTTATAATTTACCTAACCTATATCACTG
 TGCTATTAATCTATTAAATCTCCCTCAAAATTAATTAATCTCTTTTGTGTTTGT
 TGTGTTTGAGGCGAGTCTCATTTGTTGCGCAAGCTTGGAGTAAAGTGGGTGCATC
 AGCGTCACCTGATGCTTAACTGCGGAGATCANGTGGTCTTTC

GTTATCAAGTACCTCCCTGAATGGAGCTGNGTGGCTCATCTTGGCTGTCAATTCAGTATA
 TGGTAAACCCAGAGACTGATAATTTCTTCTCAGGAATGCCCACTGGAGTGTCTCT
 TTCTCTCTCTCTTTATCTTATAGAGAAATGGTAAGTGTGATCCCACTCACTCTCT
 AGTAAATCATTAATTTAGCTATAGTAACATTTCTTAAGATTTCGGCTGGCATGGTA
 GGTCAATGCCCTGTAAGCTTAGCTCTTTGGGAGGTTTAGGCGGGGAGATCACTTAAGCCCA
 AGTTCTAGACCTAGCTTTGGCAACATGGCAAACCTCGTATCTACGAAATACTAAAAAT
 AGCTTGGTAT

GTCTAGGTGTTGGCGGGGAGTGTCTTTTAGCATGCTAATGTATTATAAATTAGCGTATAGT
GAGCAGTGGAGGATTAACAGAGGTCACTCTCCTCACCATCTTGGTTTGGTGGGTTTGGC
CAGCTTCTTATTGCAACCACTTTATCAGCAAGATCTTTATGAGCTGTATCTTGTGCTG
ACTTCCATCTCATCCCCGTAACCTAAGAGTACCTAACCTCTGT?AAATTGAAGTCCAGNAG
GTCTGGGCTTATTATTAACCCAGCCCTATTCAAAATAGAGTNGTTCTTGGNCCAAACGCG
CTTGACACAGGAAAT

3640 SEQUENCE ID NO. 66

CTGCGGCGGGGGGATCTTGTAAAGTCACAGGTCAGGCTACTAAAGCGGAAAAT
 AAAGGCAAGTAAACAGCTGCGAGCAAAAAAGGGATAAAGAAAGGAATTAAGAACTAG
 CATTTTAAAGTGGGGAGGTGAATGCTTCCGAGATGGGTTATATCTCTTGGTGGG
 GCTTTCTAGCTGTTGGTAAACACCTGTCATCATCAGCATACCTGTCATCTTAATGGTC
 TCCATACCTTACTAATAGATTATACAGATGGCATCACTTAACACTTCTACTCACTCAAT
 TTGTNCAACATGGAGGTACCTCTTTTTCCTTACGCGCAAAAGCAATGGAAAGG
 TTTGTATTTTACTAGCTCCACTTCACTCAAAATTAAGCAATTTCTTTTCTTTTAAAC
 ATCCAGGACAGGTTTAACTAGGAAT

3641 SEQUENCE ID NO. 67

CTGAGCTCCAGGCACTTTTCAAATTCAGCTTCTGTGATTTCAAGCCACATATGCA
 GGAATATCTTACCTTAAATTAAGATTTAAATCTTTGTCTAGAGGCTTTGACC
 AGAGCAACTCTATCTTGAATAGGGGCTGGGTAAATTAAGGCAAGACTTACTGGCTGCA
 TTTGAGGAGGTTAGCTACTTTAGTTACGGGATGAGATAGGAAGTACGCAAGATACA
 GGTCTAAAGGATCTTGGCTCAAAACTGGTTGCAATAAAGAGCTGGTCAAAACCTACC
 AAAACCAAGATGGTGAAGAGAGTGAATCTGTGTTATCTTCACTGNTCACTATACGNTAAT
 TATTATACATTAAGCATGCTAAAGACACTCTCTTCAACACATGANTAGGTTTACAGT
 NCCATGGNAACGTTCCCGGANGTANCTTG

3642 SEQUENCE ID NO. 68

CTGNAGCTCCAGCACCAGGTTCAAGTGATTTCTCTGCGCTAGTCTGAGTATGTTGG
 GATTACAGGCAATGTCGCACTATGCGGACTTAAATTTATATTTTACTAGAGACGGGTT
 TCACTATGTTGGGCAAGGCTGGTCTCAACTCTGACCTCAAGTCACTTGGCCACCTTGGC
 CTCCCAAGTGTGTTGGGTTTCAAGGCTCTGGCTTTTACTTGAATATATCTTAAACAGG
 GGTGGATTAATCACTAGGTTTCTGGGAAGAGGTGGGCAATCCCGGAACTGAGGATCC
 CTCCCTTTTNTAGACCATCAAGGTAACTTCCGACCTTGGCATGATCTTGTAAACT
 TGTCTATGGNGTTGGGGGGAGTGTCTTT

3643 SEQUENCE ID NO. 69

CTGCAGAAAGTATGTTTCTGTATGGTATTAAGTGGATAGGGCTGAGTTATGCTGAATTGA
 ACACATAAATTTCTTTTCCACTGAGGNGCATTTGGGCGCCATGCTCTTCTGCTAGAAAT
 ATTCTTCTCTTTTCACTTGTGGATTAAATTTCTGTCTATCCCCCTCTCTTGGTGTT
 ATATATAAAGTNTTGGTGGCCCAAGAAGTAGCACTGAAATATAAATTTCTTTTAA
 TTCTCAGCAAGGNAAGTTACTTCTATATAGAAGGGTGCACTTACAGATGGAAACATGG
 CAAAGCCCACTTTGGCAAGGGAGGGGAAAGGGTTCTTATCCCTGACACACTTGGTCCC
 NGCTGNTGTGTNCTNCCCCACTZANTAGGTTAGACTGGACAGGCTTAAACTAATTTCCA
 ATTGONTAATTTAAAGACAAATATGGGGTGAATGCTTTGGGAGGAGTCAAGGAAGAGNAG
 GTACAGGTTAATTTGAATGA

0

CTTCAAGTAAATTCGACCTGGAGTCTCTTAAAGATAATGTGACATATCCGATCTTCCCTCT
GTTTCTCATCTACAGAAAACATCTTATTCGAGGTGCCATATATTCGAGCTAAAAAGAC
TTTACCTCTGACCTCTCTATATAGGATTCCTATGCAACAAGTAAAGGGCAATGATCT
TAGGGAGATCTGTCTGTATGCACTCTTAAAGGAGAACTAGTACATCTATATCT
TCTTTCTCCCTACCTCTCTTACTCTCTGAGAAATGAAGGCTAGGGCTCGAGCTTGGACC
CTGAAGTAAGCTTAGAGGTTAGAAGCTTAAAGAAAGAAAGAGGAGATTAAGTCTCTGGATGA
ACCTCGAAGCGACCTCTACTAATCTGGACTGCTACTCTCTGTACTCTCTATGAGAGAGAAA
GTATGCGCATTAAT

3664 SEQUENCE ID NO. 73

[illegible]

1663 RECEIVED IN NO. 72

TGCATCTTACGACTCTCTCTGGGTCGAAAGAGAGCTTTAAATGAGTTTACGATGTGT
 AGTTGTCTAAGTGTCTTAAAACTACTCTGGTGTCTCTAAAGNCAAAGCTTTTAGCCTTTGGG
 GAGAGCATCTAAGACAGTACGTCTAAGGGGAGGGGTAGAGTCAGAGGGCATTAAGGAATA
 TATCCATATAGGTTAGGTTAGGTGTGGCAAGCTTTATAGAACAAACATTCGTAAGCTA
 CAGACACAGGCGATCTCTCTCTACGCTTCTCAAGGTCTATTAACAAAGTAAATCA
 CAAATGCTGAACCAAC

3570 SEQUENCE ID NO. 72

[illegible]

CACTCAATCTGTAAGCTGCGCATCTGATGAGTGTGATGTTTGTGATGATCTAATCTCTCTG
 GGGTCTCTGTAATAAATACTAGGTTGCTATAGGTTGAGCATAGACACACGATCTTTGATTTCC
 ACCGTTCTCTCTCTCTATGTTGGGTAATCTTCTGACGTAATCTAACTGACCTAGGAAGA
 AAAAAATGTTTGTCTCTAGACATAGTAAATTTTATGTTCTGCTCTCTCACTGTTGGA
 ACATTCAAAAAATACAAAAAGGAAGCTAGGTTGATTTGTAATGCCAGGCTCAGAGGCTGA
 GGCAAGGAGCATTCCTTGGGCTGAGGATTTCAACAGCAGCTTGGGCAACCTAGCAGACCT
 TGCCTCTATTAAGAAAAACAAAAACAAATATTGGAAATATTTACATGCAATGGAACTCA
 TATGTAAGAAAAAATAGTGAAA

CGTGTATTATTAAGTGAACCGCCAGGAGGATAGTCTATGACTACAATGACCTCGATCATGAT
GGCAGCAATTCAGAAATTGATGCGAGGGCTCTCTGGGCCACAGTCTCGGTATGCTTCTGTGAA
TGGGGTATAGATCTTACAATAAAAAACAACACAAAGCCCTTAGGTCAAGTGTCTAAATGGAGAT
CACCACACCACTTACCACTCTCAACACAGAAATTTCTTTCTTAAATCTCAATCGATATCT
TATAAGTCACTTTTCCCAAGCTCAACAAATCTAGCAAGAAATTTTAACTCGAGAGAAAAC
AGCTACACTCTTAAAAATTCCTTCAAGAAAAATGTTCAACATATGGAAAGAAAGGACTTAACT
TGTGAAGCAGACACTGGCTCAATCTAGTGGGTGCTTATATGAATAAATATTAACACT
CATCAAAATTTTGTGGGTACAGTTATATAGGAACCTGGTATGGAAACAGATTTCTGCCACA
GAAAACACGAGGGT

[illegible]

AGGA?CGCTTGGGCCCAGGAGTTTCAAGCAGCTTGGGCAACGTAGCAAGACCCTGCCCTC
TATTAAAGAAAACAAAAACAATATTGGAAGTATTATATGCAATGGAATCTATATGTG
ATGAAAAAATTAGTGTAAAATATATATATGATTAGNTATCAAGATTTAGTGATAAT
TGTGTTAT?TGGGATTTCAGTGCCTTTTAGGCCATTGTCTCAAAAAATAAAAGCAGAA
AACAAAAAAGTTGTAACTGAAAAATAAATTTTCATATATAGCACAATCTAAGTGG
TTTTCGTTGTTGTTGTTGTTGTTGTTGAAGCAGGGCCTTGCCTCTCCACCAGGNTGGAGTG
AAGTGCAG

1. The first group of people who are interested in the study of the history of the United States are the people who are interested in the history of the United States.

3744 SEQUENCE ID NO. 82

3745 SEQUENCE IN NO. 83

3749 SEQUENCE IS NO. 94

1.35

3750 SEQUENCE ID NO. 85

CTGCAGTGAAGCAAAATCGTGCCTCTCACTTCACTCCAGGCTGGGTGACAGGGCAGGC
CTTGCCTTCAACCAACCAACCAACCAACCACTTCTAGATTGTGCTATATATGAA
ATGTTTATTTTTCAGTTACACCTTTTGTGTTTCTGCTTTTATTTGTTGAGACATGGC
CTAAAAGGCAATGAAATTCAAAATACATAAATATCTAAATCTTGAATACATTC
ACATATATATATTTTACACTAATTTTCTAGACATAAGATTCCATGCTATAAAAT
CTTCCATATTTGTTTGTGTTTAAATAGAGGCAAGGCTCTTCTACTTCTCCCTAG
CTGCTTGTGAATCTCTGGGCAAGCACTCTCTGCTCTGAGCTCTGAGGCTGGCATTA
CACATGCACTGGCTCTCTTTTGT?TTTTTGAATGTTCCACAGTGAGGAGGAAGAAAA
CTTAAAATTAACCTATCTT

3751 SEQUENCE ID NO. 86

CTGCAGATGAGAGGCACTAATATAAGCCATATTAACCTTTCTTCTGACAAACCTTGTCA
GCCCACGTGGTTCTGTGGCAGAACTGGTTCTATAACAAGTTCCTAATAGCTGTAGCC
AAAAAACTTGAATGAGTATATATATTTCTAATATAAGCACTCTACTAGATTGAGCC
GTGTCTGCTTCAATGTTAAGTCTTCTTCCATATGTTAGACATTTCTTTGAAGCAAT
TTTAGAGTGTAGGCTTTTCTCAGGTTAAAAATCTTAGGATAGGATGGTGAATGGGG
AAAAGTGACTATAAGATACCAATGAAATTAAGAAAAAGAAATCTTGTGTTGGAGGTTG
TAATGTGGGTGGTGAATCTTCTTAACCTGANTAGGG?TTTGGGTTTGGTTATTTGA
GATCTATACCCATTCAATAGAAGATCCG

3752 SEQUENCE ID NO. 87

CTGCAGCCAGTAAACAGTTTCTTCCAGGTTCACTATTTAAGGTACATCTCAAAGCAT
AAGTACAAATGTTGTTTGTGATGACAGAGCAATGAGCAGCCAGACCTGAGTAA
AAGAGAGGAGCTGAATGCTTCTGTACATCTATAGATATCTAGACTTTTAACTGCA
TACTCTCTGTACATGTTTCTTCTCTCAGGTTTCTAGAACTCATAGTACCGGTTCTG
TTGTTAATCCAGGTTCAACCTTACCTTGATCTGCTGAGATCTGATTTACTGAAAT
GTTTTTCTTGTGCTTATAGAATGACATAGAGAACGGCAGGAGCACACGACAGACGGAG
CCTTGGCCACCTTGAGCAATCTAATGGACACCCAGGTTAACTCCCTGAGGTGGTG
GAGCAAGATGAGGAAGAGATGAGGAGCTGCATGAAATATGGCGGC?AGCATGTGATC
ATGCTCTTGGCCCTTGTGA?TC

3753 SEQUENCE ID NO. 88

CTGCAGTGTCTCTTTTCTCCACTTAAAACATGAAGTAATAACCCCTCCTTCTCTCACTG
CTCCTGACCATCACAGAGGATGGGCTCTCCAGCTAAGTCATGCCCCCTTCAATGACAGGC
CTTCAAGGTGATGATGACATGCTGTAAAGAAAAGCCACACTGGTTTGGAGATTAATAAA
CAAAACTCATACGTACAGCTGCCCATCTTCCGGGTATAAAGCTGACTGACTTAATGGT
AGCCACGACACCACTGTCAGAGAGTCAAGGGACAAAGAGCATGATCACTGCTTGGC
GNCATATTCAATGTCACTTCTCATCTTCTCTCATCTTCTTCCACCTGCTGCGGGA
GTTACC?TGGGTCTCCATTAGATAATGGGTGAGGTTGGCCAGGCTCCCTCTGTCTT
TGCTCTGCGCTTCTCTATTTCTATTCTATAAGCACAGAAAAACATTT?CAGTAAATCA
GAT?CTCAGCAGAATCAAG

3754 SEQUENCE ID NO. 88

TAACTCCGAGGTTCAAGATTTCTTCTGCTTAGGCTGCTTAGTAGCTGGGACTATAGGT
ATGTCGCACTATTCTGAAAACATACAGTTTGAAGGTAGTGTCTGGCTGGGCSGAG
TGGTTCACTCTCTTATCTGAGCACTTGGGAGGNCAGGTTGGCGGATCTCTGAGGT
AGGAGTTGAGACGAGCTGACCAACATGGGATAAGACTGATCTCTACTAAAAATACAA
AAAAATTAGCGAGGCACTGGTGGNGCATGCTTAAATCTCAGCTACTCAGGAGGNTGAGGNA
CGAGAAATTGGTTGGAACTTAGGAAGCAGAGGCTGTGGTGGAGCCGAGATCTGACCATTTGG
ACTGCAAGCTGGGNAACAAGAGTGAAGATCTTTCTTAAAAAAGGCTAGNG
TTTTGNCCTGGNGCGGGGGTCTGCTCTTAACTCCAGNATTTGGGANGGCGAGGNGGGGG
GTCANRANGNNAAGTAGTCCG

3755 SEQUENCE ID NO. 90

GAAATTCCTGCTGACATGTCCTATGTTCTTTCTCCCTACTGCTTCTACTGTGAGTAAATG
AAGGGTAGGGGCTCCAGCCTGGACCTGAAAGTAAAGCTAGAGGTTAGAAGCTAAAGAGAAA
GAAGGAGATTGAGTCTTTGATGAACCTGAAGCCAGCGTACTAATCTGACTGCTACTCT
CTGCACTACTCTATGAGAGAGAAAGTATTTGCAATATTTAAACAGTTGGGTTGATTTTC
TATTAAACAAAGTCTGAAGACATCTCTTAAAGCCAGACTGAATATTTTAAAGCTCTATGG
GTCAATATGCTCTCAGGGCAACACTCTACTGTGCTACTGTAGTGTGAAGCAGGCAAG
ACAAATGTAATTAACCAAGGAGGCTGGTCACTTCTCAATGAAGTTTATCTCAAAATGGNGA
ATCTTTGGTATTAACACCTNGGGGCAAGGTAGGAGAGATCTTGGCTGTGGTTGTNGNTGG
CAATGTTGGTCTTTTATACNG

3757 SEQUENCE ID NO. 91

GAAATTCCTGCTGAGAGTTCCATACACACACTCTCTGAGAGTCAATGGCTTACT
TGTCTCATAGCTCTCTAAATATAGGGAGTCAAGAGTAAAGTCTGCTTGGCTGGGAA
ACTGGCACCTGGAAATAAAATGTTTCTGCTGAATGAGAAACAGGGGAGAGATGGATATG
TGACATTAATTAAGCACTCTGAGTTGCAATTACTCTGAGATGAGAGGCACTAATTAT
AAGCCATATTAACCTTTCTCTGACCAACCTTGTCTAGCCCTCTGGTTTCTGTGGCAGAA
TCTGGTTCTATAACAGTTCTTAATAGCTGTAGCCAAAAAATTTGATGAGGTAATTA
ATTATTTCAATATAAGCACCTACTAGAGGAGGCTGTCTGCTTCACTGTAAAGTCC
TTCTTTCCATATGTTAGACATTTCTTTGAAGCAATTTAGAGTGTAGCTGTTTCTCTA
GGTTAAAAATTTCTTAG?TAG

3810 SEQUENCE ID NO. 92

TATGCTTGCCTATTCTTGTCAAGTAACATACTGATCTAGGAAAGAAAAATGTTTGT
CTTCTAGAGATAAGTTAAATTTAGTTTCTTCTCTCTCACTGTGGAACATTCAAAAAAT
ACAAAAAGGAAGCCAGGTGCACTGTATAAGCCAGGCTCAGAGGCTCAGGCAAGGAGATCG
CTTGGGCCCCAGGAGTTCAAGCAGCTTGGCAACGTAGCAAGACCTGCTCTATTAAA
GAAAAACAAAAACAAATATTGGAAGTATTTATATGCAATGAACTCTATGTCTATGAAAA
AATTAGTGTAAAAATATATATATATGATTAGTTATCAAGATTTAGTATATTTATGTTA
TTTTGGGATTTCAATGCCCTTTTAGGCCATTGCTCTAAAAAATAAAAGCAGGAAAAACAA
AAAAAGTTGTAACTTGAATAATAACATTTCCATATTTATAGCCAACTAAGTGGGTTTNG
GGTNGGTTGGGTTGGTTGGT

TATCTTTAACAGGCTCCCAACCCCTTAAAGGTACGATTTTCTTCTTGTGGAGA
 CAGGGTGTCACTTGGTCCGCTCAGACTGGAGTGCAGTGGCACCATCTCAGTTCAACCAAC
 CTGGCTCTCTGGTCTCAGCACTCTCTGCTTAAAGCTCTCAGTAGGCTGAACACCG
 CTTGGCGCCCTACCAACGCTAGGCTTATCTGGCTCTTAACTAGAGACAAGGCTTCTGCA
 GTTGGCCAGGCTGGTCTCAATCTCAGACCTCAAGTGATCCGCTCCGCTCAGACTCCAA
 AGTGTGAGCACTACAGTGTGTACCACTATGCTCCGTCGCTCATCTCTCTTAAACAT
 CCAACCTTACCTAGTCTCCTG

[illegible]

GGTGGGCGTTCTAGACTAGTGGCICCCAGAGTAGAAGAAGTTTCTTAGTACAGAACTA
AATGAAANGTCTCCCATGTCTACTCTTTCTACACAGCACGCGATCCATCCGTTTTCT
GANTCTTCCGTCACCTTTCCGCTCTTTCTATTCACAAAGCCGNCATGTGTCACTCTGGC
CCTTCTCTAATAGCGGTGTATNTACAGCTCCAGACGCGCTGGTGGNCGGTCTAGGCGGC
TCTCACTTCCAGTAGGGGTGGCCGNCAGGNGTTCCTCCCTCAC TCCCGCGGCGGGGTG
GTNGTCCGNCGCGGCTGACAGTCCCGACGGCTCCGCTCTCTTCTACTGGCGGTCTG
TATATTCATATATATATAGCA

GGAATCCAAAGGAAGTTAGAGGGCCAGCTCAGTCTACACCTGCTACTGNTCAGTGGCCACCC
 GGTCAAGGGGAGACCAACACATGGTAAAGGTCAGGGGCTTCTTGGAAAGGCAGTCAGCAGCC
 TGTGCAGAAGATTTCTCCACATCTGTCAGTTAAAGGGGAGCTGGGGGCAGGAACCTCAGCTG
 GATCTCTGCTTCCACAGTGTCCAGGGGTTCACAATCTGTGTTACTCGTACGGATATTT
 AATCTTGAGGGGTGCTACAAATGGGATCTATCTCTAATATGATGTTTCTTACTAGTT
 GTTATGGAAGGTTATTGATGGGTTTACCTTTATTTNGTGGATGGAATGATCCTC

AATGTCACGGATTCCTTATGGTAGTACACCCATCAACCTTTTGTAGAAATAAAATGAATT
 GAGAGTGTATACAGTCTAATTCTATATCAGTGTAACTTTATTGGATATATCAGTAATA
 GTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNGGNGANAGAGTCTCGCTCTGT
 CGCCAGGTTGGAGTGNAAATGGTGCCATG

3838 SEQUENCE ID NO. 98

AACAGAGGTTTCTGGGTGGGGTGAATAACCGGGGGTGGATATTTGTTGCGAATAC
 CCGCTGACCGTAAACGTGGCTTATGGGCAGCTGGCTGACTTGGTTCTATGGCGGT
 TTGTGCTGGGTGCGGGCTGGTGGTAAATTTGACGATTTGTGGGCAAGCGAATTTCT
 TCGATTGGGGCTGGCTATTCCTTCTTATGCTCTGGCTTAGGGATTATCGGGCTTT
 ACCTGGCGCATGGCTGGAGGAGCTCGGGCGTTCCAGCAGNATGTGATTAAGTGGAAC
 AGGCGGACCGTGAAGGTTTGGAGGATGGCGGAAAGTTCCTTAAAGAGATTGGCATT
 AATACTGGNCCAGNCTGTTGAATGTTGGGCTTGGTAAATGGCAACCAACGTGATTACT
 AATGTTGTTGACCAATAATGGCTAGTATTTGGCGGATAACCTGAATTATC

3841 SEQUENCE ID NO. 99

TAAATTAATTAAGGCTTCTCTAGGTCATCCATGNCGGNTTATTACATCATCTCT
 ATTGNTGNTGCTCTTTTACATTCATTACTTGGGGTAAAGTTGTGAAATTTGGGGTCTG
 TCTTTCAGAAATTAAGTACGTTGGTCTGTGTAGCTATCATTTAAAGCCATGTACTTTGNT
 GATGAATTAATCTGAACTTTAAATGNTCCACATTAAGGTCATCTTGGTATATAAAG
 ACTAGNCAGTATTACTAATGAGCATTCTTCTGTTGCTCTTCTGCTTACATTAAGTAGAA
 CTGAAGTAAGTTAAGACTACGTTAATTTAAGCTTTGGGGAAAGGATTATAGGCTT
 CTAGTAGGAATCTTGTGCTATCAGAACTTTTAAAGAAAGGGTNTCAAGGAATNGTAT
 AATACCAAAATTAATGAT

3885 SEQUENCE ID NO. 100

AAAAAAGGCTCTTGGAGTTCGAAAGGGAAGAAACAGCACTTTGTGCTTACAA
 TTACTACTGTTATAAAAACACTTCATAGATTAATTAAGCAGAAACAACTTTCTTTC
 ATGTTGTTCTCTTGGGCGAGCTGTCTAGGACCTCAAGGCTTGTGCTCTTGCAGGC
 TCCGATTAAGTCTGAATAGGATTCACAGGTCATCCATTTGTAGTTAAATCTTAGT
 AAATCAATTAAACCAATCAATTAATTAACACCAATTTGTAACCGGAATGATTAAT
 TTTGTCACTACTTCACTACTCAAAATTAATTAAGCTCAAGGTAGGCTCTTTAG
 CATTAAGTAGGAACCTCTGAGCAGCAACCAATTTCACTCAAAAGATTAATTAAT
 GACTTTCTGATTTCAGCAAAAGGGGG

3886 SEQUENCE ID NO. 101

GGATCGGCGCTCTCGGCTCCAAAGTGTGGGATTACAGGCTGAGGCACCCACCTG
 GCTTTTTTTTTTTTTTTTGGNGAGACAGAGTCTTACTCTGTTGCCCAAGCTGGAGT
 GCAGTGGTGCATCTTGGTTCACTGTAACCTCCACCTCCAGAGTTCAAGCAATTTCTCTGC
 CTCAGTTTCTGGAGTAGCTGGATTACAGGTGCTTGCCTCAGCCCTGGCTAAATTTGGN
 ATTTTTTTTTACTAGAGACAGGTTTCAACCAATTTGGGCAAGGCTGGTCTTGAATCTCTGA
 CCTTGTGATCCACAGCTCGGCTCCCAATTTGTTGGGATTACAGGCTGAGCCACAC
 AACCAGGCTAAAGTTTTAAACATGCCAGTGTATTACATTAATGCCATAGCAATTAATG
 ACATA

1387 SEQUENCE ID NO. 102

GGATCCGCGCGCGCTTGGCGCTCCGAAAGTGGCTGGGATTACAGGCGATGAGCGACCGCTCCTG
GCTGAGTCTGCGATTTCTTCCAGCTCTACCGAGTTGTGTCTTAAAGCAAGTCACTGA
ACTTCTCTGGATTCCCTTCTCTTTTGTAAATAAGCATGTTATCTGTCCNCTGCTGCTT
GGGCATTGTGATTAAGGATAAGATGACATATAGAAATNGCAAATTAAGCGCTAGAC
AAATGATTTTATTAATAATAAAGATTAGTTTACTTTGGGCCAGCATAGAAAGGAAT
GTTGAGAACTTTCTTAAAGGACTTCTCAAGCTCCCTTTGGTGTATATCACTGTGTCATTA
CTTACCTTNGGGCTGAAAAATGTT

1388 SEQUENCE ID NO. 103

GAAAAGGGAAAGAAAAACAGAACTTTGTGCACTACAAATTATACTGTTATAAAAAACACTT
CCTTAGATTACCTTAAGCGAAACAACTTTCTTTCTATGTGTTCTCTCCAGGCGCAAGC
TGTCTAAGGACCTCAAGGCTGTTGTCTACTTGCAGGCTCCAGATTAGGTCTGAATAGG
ATTTCAACAGGTCTCCATGTTAGTTAAATCTTAGTAAATNCA

1389 SEQUENCE ID NO. 104

GGATCCGCGCGCTTGGCGCTCCGAAAGTGGCTGGGATTACAGGCGATGAGCGACCGCTCCTG
GCTTTTTTTTTTTTTTTTTTGGGAGAGACAGAGTCTTACTCTGTTGCCAAGCTGGAGT
GCAGTGGTGCATCTTGGTCTGCTGCACTCTCAGCTCTCAGAGTTGAGCAATTTCTCTGC
CTCAGTTCTGGAGTAGCTGGATCTCAGGTTGCTGCCATCAGCTCTGNTAAATTTGG
ATTTTTTTTAGTAGAGACAGGCTTTTANCATGTTGGCCAGGNTGGTCTTGGACTCTCTA
CTGCTGAACTCCAGGCTCCGCTCTAAATTTGCTTGGATTAAGGGGTTAANCAC
CACACCTCAGCTTAAAGTTTAAACATNCAAGTCTTTAAATNATGNGATACCAAT
TATTTGACAAATTAATTTAT

1391 SEQUENCE ID NO. 105

GTCTTCCGCTTCTCTCAGAGTTTGTGCTTACATTAATCTCTGCTTCTTCAA
GAAAGCAATGTGAGCTCTTCCATCTCCATCAGCTTTGGGCTTGTCTTCTACTTTGCCAC
AGATTATCTTGTACAGCTTTTATGGACCAATTAGCATTCATCAATTTATATCTAGCA
TATTTGGGCTTGAATCTCATGATGTTTCTTCTTGAATATAACAAATCTGGGGAGGA
CAAAGGTGATTTCTGTGTCTCATCTAACAAAGTCAAGATCCCGGCTGGACTTTTGG
AGGTTCTTCCAGTCTTCTTACCACTTGCATATGGACTTTGGAAAGGAGGTGCCT
ATAGAAACCAATTTGGAACTATCTTCTCGGAGGGGACTGTGTCCCTCGGTGGCAGAA
NCTACCAAGATTTGCGGCTCAGGTCAA

1398 SEQUENCE ID NO. 106

GGATCCGCGCGCTTGGCGCTCCGAAAGTGGCTGGGATTACAGGCGATGAGCGACCGCTCCTG
GCTGAGTCTGCGATTTCTTCCAGCTCTACCGAGTTGTGTCTTAAAGCAAGTCACTGA
ACTTCTCTGGATTTCCCTTCTCTTAAATAAGNATGTTATCTGCTGCGCTGCTT
GGNATTTGAGATAAGGAT

1393 SEQUENCE ID NO. 107

CTGCAGTGAAGGCTGATTGCACTTACTTCAGCTTGGGCAACAAAAGAGACCTGG
CTCAAAAACAAAACAAAACAAAACAGGTAATTAATTTAAAGGGAAGTATTAATTA
AATAATAGCAAGTTATATAGGTAATGTAATTAATAAGGTGGGATTTAATATCTA
ATGTTTGGGAGCCATCATTATTTAAATATCTTTGGTCAAAAATTAATGATCTT
TAAAAATTTGTATATTTTTCAGGGAAGTGTTAAAAACCTAATAACGTTTCTTGGAC
TACATTACTGTTGCACTCTCTGATCTGAAATTTTGGGTGTTGGTGGGAATGATTTCTATTCA
CTGGAAAGGTTCACTTCTGACTCTGAGCAGGCATATCTCATATGATTAGTGGCTCATGATC
CTGGTCTTATCAAAAGTACTCTCTGAAATGCACTGCTTGGGTCACTCTTGGTGTGATTCAT
GTATATGGTAAAAACCAAGA

1394 SEQUENCE ID NO. 108

CTGCAGCTTGAACCTCTCTGGATCATCTGATCTCTCTCACTCAGCCTCTAAGTAGCTGG
AACTACAGGTGTGCACTCACTCTGCTGGCTAAATGATTTTCTGTAGATACAGGTTT
TNGCAGTGTTCCTCAGGCTGGTCTGAACTCTGGGCTTAGGTGATCTGCTGGCTCAGCC
TCTCAAGTGTAAAGATTACAGGCTAGCTACCATGCTCAGCCGAAATCTTCAATGAA
AAAGTTACTATAGCTAATTAATGATTACTGAAGAGTTATGGGATGATACCTTACCAT
TTCTCTAAATCAGATAAAGAGATGAGGAAAGAAAACCTCTCAGTGGGGCTCTCTGTA
CAAAACAAATTAATCAGTCTTCTGGGTCTAACATATACTGAAATCAAGGCAAGATGAGCC
ACCTAGTCTCTCAGGGAGGTACTGGATACACCTAGGGCTGAGGGACTAATCATATG
AGATATGCTGCTGAGTCTCA

1378 SEQUENCE ID NO. 109

CICNN
ATATGANTTTCTCAGGCTCTTAATTTCAAACTTTGATTCACAGTCTCTTTAAGGCTG
TTCTGTTTAACTCCAGGTTGGTTAAATATTCAGCTACCTGAGGAGCTTTGATTAAT
GGACCTCAGTTAGTAGTTTCTACCTGGCAGCATAGAAATCACTTGGGAGCTTTTA
A

1379 SEQUENCE ID NO. 110

TNN
GGTCACTCTTNGGTATCAATTCACAACTCAAACTATCACTGCTCTACTCAACCCACAA
TATATGAGAGAAGTCAGTAATGATACAAATTAGGCTTCAGCTGTGTTTCTTTCTT
TNGGGGTTTNTACAAACAGGTTCCAGATTCTATGTGACTGACTCTGGAGTCTTAATCT
T

1390 SEQUENCE ID NO. 111

NNCCCTCCTCCTNN
CGGTGGCCGTCCTCTAGAACTAGTGGATCCCCCGGNTGCAGGACCTCAACGCTGCCCGA
GATGCTCCCGCTGCGGTTGCTGAGATGSCGGACCTGATGGATATGTTCTGCCAAGGTT
GGTTTGGCATTACAGTTCTCCGCAAGAAATGATTGGCTCCAAATCTTGGAGTGGTGAA
T

66

I G

4

II G

G

五

T

11

A

3

1

22

TTTGGCTGGTACTGTTTGAAGCAATGGTGAATCTGGGCGCTACGCTGGCGGGTGTGTG
 CTCTGCGAGGGGGGACGACAGCAATTGGGGCGAAGTCCGCACTCTCTTGTGCGAGGCGGAA
 GACGGGCTGTGATGCTCTCTGTTGTCTGGGACTGTCTCGAGGCATGCCAGTCTCAGTGAC
 TCTGTGTCTGGCGCTCTCTCTGCTTCTCAATCTTCTCAGCTTGTGGTCAGCTCTGCTT

ATTGCTCGTANCCGCAATACATATAGGGGCACTTGGAGCTTCACTGGCGGTGGCGGNCCTCG
 GCAGCGGNCCTGGNCCTTTGTGGCCCCCGGGCCGGCAAGCGGGTGTCTCTAAAAGA TGAGGG
 GCGGGGGCGCGGNCGGTCTGGGGCTGGGGAACCCCGTGTGGGAACCAAGGAGGGTGGCGCCCG
 TTTCTCGGGCTTCGGGCGCGGGCGGGTGGAGAGAGATTCCGGGGAGCCCTGGTCCGGAAA
 T

ATATGCGAGCCGCGTAACTTAACCTCTCTAAAGGGGAACTAAAGCTGGGTACGGGGGCCGGGGC
TGGAGTTCGACGGTATCGATAAGCTTGACATCGAATCTGTCAGCCCGGGGGATCGGGCC
CGGGGCTTCCAAAGTGTGGGATACAGGCGTGAAGCCACCGCCCGCGGCTCTCACTTT
CACTTCTATTGGCTAGCGCTGCTCTAAATCTTCTGTTCCTTCTGCAACACAGGGCTTAACTA
C

[illegible]

885-886.885 SEQUENCE ID NO. 122

CTTAGTCTGTCTCTAGTCTACTTAACTTAAGTCTACCTAAATAGAACTGTCTCAGAGC
TCTTAATGTCAAACTTTGATTACACAGTCCCTTAAGGCACTCTGTCTTAACCCGAGG
TGGTAAATACTTCCAGCTATCTAGGAGCTTTTNGATAACTGGACCTCAGCTTAGTACT
TCTCTACCCCTCCCAACATTAAGATCACTTGGGAGCTTTAAAACTGTAAAGCTCTGCC
TGAGATACTCTACTCAATTAATTGTGTAGTTTTTAAATTCGCCAGGAAATCTGGTA
TTCTGTCTAGGAACCTCTCCCTCAGCTTAGCAATACAGATATGTAGGAAATTAGCTCT
GTAAGCTTGGTCTTACAGCCGATAAACAGATCTCTTACNCCCTGGGACTTAATCACT
GAGAGTTTGGGTGGNGCTTNGNATTTAATGAC

885-886.886 SEQUENCE ID NO. 123

GACACCACTTCCACATTAATTATGAAGCAATTTCCAGCCAAACTCAATCAAGTCTGG
GTTTTAACTAGTTACTGAATAATCCCTTGGGGGCTTAAATTTAGAACAGAGCTTC
ATYCAATCTGGAGAGAGAGCTATGAAAAAAACCTAGCTTGGGTNGGTTTCAAGGGTNC
TATGACACATTTGTACTTTATCTTTAATVCTAGTAAAGAAATAGAACTCTGAAATTA
GTAAACTACTTGGAAAAAATTAAGGATACAGAAATTTCTATCTTAAATGATGTCTGG
GCCTCTGTCAATTTAGTNGGNTGGTTAAANCCAGAGGCTAAGAGNATCTCTATCTCT
GTGNGGGG

885.INT SEQUENCE ID NO. 124

GCTCATCTGTTTCTCCGGGAGGCTGTGGGGGAAGAACTCTCCACACAGNATAAGAA
TCTCCCTCCACAGGATAGAGAACTCCCTCCACAGCAAGAGAGCCCCCTCAGGCTA
GAGAACTCCCTCCACAGGATAGAGAGGCTCCCTCCACAGCAAGAGAACTCTCTCACT
CTGGGTTTTTACCGCCAACTAAATCTCAGAGGCTCAGCATCTATTAAGATAGAAA
TTCTGTACTTTTAAATTTTCTAGTAGTTTACTTATTTTCAAGATCTATTTCTT
ACTAGAACTAAGGGAATAAATTAACATGTGTGCAATTAACCTATGAAACCTACMAA
GCTAGGTTTTTCTAGCTCTCTCTCAGATTGAATGAACCTCTGTTCTAAATTTAAC
CCCCCAGGGAAATATTCACTTAACTATGTTAAAAACCTCAGACTTGTGATTGATTTTGGC
TGAAATGCTTTCTAATTAATGTGTGATGTGCTC

894.915.GEN SEQUENCE ID NO. 125

GTATAATGCAGGTGCTATAAGGTGAGCATGAGACACAGATCTTTGCTTTCCACCTGTTC
TTCTTATGGTTGGTATTCTTGTCACTAACTTAACTGATCTAGGAAAGAAAAATGTT

Primers

849	5'-atctccggcaggcatatct-3'	Sequence ID No: 126
892	5'-tgaaatcacagccaagatgag-3'	Sequence ID No: 127
885	5'-tggagactggaacacaac-3'	Sequence ID No: 128
893	5'-gtgtggccagggtagagaact-3'	Sequence ID No: 129
890	5'-ccatagcctGtttcgtagc-3'	Sequence ID No: 130
891	5'-ccatagcctAtttcgtagc-3'	Sequence ID No: 131

SEQUENCE ID NO. 132

FILE NAME: ARMP.UPD

TGGGACAGGCAGCTCCGGGGTCCGCGGTTTCACATCGGAAACAAAACAGCGGCTGGTCTGGAAGG
AACCTGAGCTACGAGCCGCGGCGGCAGCGGGGCGGCGGGGAAGCGTATACCTAATCTGGGAGCCT
GCAAGTGACAACAGCCTTTGCGGTCTTAGACAGCTTGGCCTGGAGGAGAACAACATGAAAGAAAG
AACCTCAAGAGGCTTTGTTTTCTGTGAAACAGTATTTCTATACAGTTGCTCCAATGACAGAGTTA
CCTGCACCGTTGTCTACTTCCAGAATGCACAGATGTCTGAGGACAACCACCTGAGCAATACTGT
ACGTAGCCAGAATGACAATAGAGAACGGCAGGAGCACAAACGACAGACGGAGCCTTGGCCACCCTG
AGCCATTATCTAATGGACGACCCAGGGTAATCCCGGCAGGTGGTGGAGCAAGATGAGGAAGAA
GATGAGGAGCTGACATTGAAATATGGCGCCAAGCATGTGATCATGCTCTTTGTCCCTGTGACTCT
CTGCATGGTGGTGGTCTGGCTACCATTAAAGTCAGTCAGCTTTTATACCCGGAAGGATGGGCAGC
TAATCTATACCCCATTCACAGAAGATACCGAGACTGTGGGCCAGAGAGCCCTGCACTCAATTCTG
AATGCTGCCATCATGATCAGTGTCTTGTGTCTGACTATCCTCCTGGTGGTTCTGTATAAATA
CAGGTGCTATAAGGTCTATCCATGCCTGGCTTATTATATCATCTCTATTGTTGCTGTTCTTTTTTT
CATTCAATTACTTGGGGGAAGTGTTTAAACCTATAACGTTGCTGTGGACTACATTACTGTTGCA
CTCCTGATCTGGAATTTTGGTGTGGTGGGAATGATTTCCATTCACTGGAAGGTCCACTTCGACT
CCAGCAGGCATATCTCATTATGATTAGTGCCCTCATGGCCCTGGTGTATCAAGTACCTCCCTG
AATGGACTGCGTGGCTCATCTTGGCTGTGATTTCACTATATGATTTAGTGGCTGTTTTGTGTCCG
AAAGGTCCACTTCGTATGCTGGTTGAAACAGCTCAGGAGAGAAATGAAACGCTTTTTCCAGCTCT
CATTTACTCCTCAACAATGGTGTGGTTGGTGAATATGGCAGAAGGAGACCCGGAAGCTCAAAGGA
GAGTATCCAAAAATCCAAGTATAATGCAGAAAGCACAGAAAGGGAGTCACAAGACACTGTTGCA
GAGAATGATGATGGCGGGTTTCAGTGAGGAATGGGAAGCCAGAGGGACAGTCATCTAGGGCCTCA
TCGCTCTACACCTGAGTCACGAGCTGCTGTCCAGGAACCTTTCCAGCAGTATCCTCGCTGGTGAAG
ACCCAGAGGAAAGGGGAGTAAAACCTTGGATTGGGAGATTTTCAATTTCTACAGTGTCTGGTTGGT
AAAGCCTCAGCAACAGCCAGTGGAGACTGGAACACAACCATAGCCTGTTTCGTAGCCATATTAAT
TGGTTTGTGCCTTACATTATTACTCCTTGCCATTTTCAAGAAAGCATTGCCAGCTCTTCCAATCT
CCATCACCTTTGGGCTTGTCTTCTACTTTGCCACAGATTATCTTGTACAGCCTTTTATGGACCAA
TTAGCATTCCATCAATTTTATATCTAGCATATTTGCGTTAGAATCCCATGGATGTTTCTTCTTT
GACTATAACCAAATCTGGGGAGGACAAAGGTGATTTTCTGTGTCCACATCTAACAAAGTCAAGA
TTCCCGGCTGGACTTTTGCAGCTTCCTTCCAAGTCTTCTGACCACCTTGCACTATTGGACTTTG
GAAGGAGGTGCCTATAGAAAAAGATTTTGAACATACTTCATCGCAGTGGACTGTGTCCCTCGGTG
CAGAAACTACCAGATTTGAGGGACGAGGTCAAGGAGATATGATAGGCCCGGAAGTTGCTGTGCC
CATCAGCAGCTTGACGCGTGGTCAACAGGACGATTTCACTGACACTGCGAACTCTCAGGACTACCG
GTTACCAAGAGGTTAGGTGAAGTGGTTTAAACCAAACGGAACCTCTTCATCTTAAACTACAGTTG
AAAATCAACCCAATAATTCTGTATTAAGTGAATTCTGAACCTTTTCAAGAGGTACTGTGAGGAAGA
GCAGGCACCAGCAGCAGAATGGGGAATGGAGAGGTGGGCAGGGTTCCAGCTTCCCTTTGATTTT
TTGCTGCAGACTCATCTTTTAAATGAGACTTGTCTTCCCTCTCTTTGAGTCAAGTCAAATAT
GTAGATTGCCCTTTGGCAATTCTTCTTCAAGCACTGACACTCATTACCGTCTGTGATTGCCATT
TCTTCCCAAGGCCAGTCTGAACCTGAGGTGCTTTATCCTAAAAGTTTAACTCAGGTTCAAA
TTCAGTAAATTTTGGAAACAGTACAGCTATTTCTCATCAATTCTCTATCATGTTGAAGTCAAAT
TGGATTTTCCACCAAATCTGAATTTGTAGACATACTTGTACGCTCACTTGCCCCCAGATGCCTC
CTCTGTCTCTCATTCTCTCTCCACACAAGCAGTCTTTTCTACAGCCAGTAAGGCAGCTCTGTC
RTGGTATCAGATGGTCCCATTATCTAGGGTCTTACTCTTGTATGATGAAAAGAAATGTTGTTATG
AATCGGTGCTGTGAGCCCTGCTGTGAGCCTTCTTCCACAGCAAATGAGATGTATGCCCAAAGCG
GTAGAATTAAAGAAAGAGTAAAATGGCTGTTGAAGCAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQUENCE ID NO. 133

FILE NAME: ARMP.PRO

MTelpAPLSYFQNAQMSedNHLSNTVRSQNDNRERQEHNDRRSLGHPEPLSNGRSPQGNSRQVVEQ
DEEDEELTLKYGAKHVIMLFVPVTLCMVVVVATIKSVSFYTRKDGQLIYTPFTEDTETVGQRAL
HSILNAAIMISVIVVMTILLVVLKYRCYKVIHAWLIISLLLLLFFFSFIYLGVEVFKTYNVAVDY
ITVALLIWNFGVVGMISIHWKGPRLQQAYLIMISALMALVFIKYLPewTAWLILAVISVYDLVA
VLCpKGPLRMLVETAQERNETLFPALIYSSTMVWLVNMAEGDPEAQRRVSKNSKYNAESTERESQ
DTVAENDOGGFSZEWEAQRDShLGPHRSTPESRAAVQELSSSILAGEDPEERGvKLGldFIFYS
VLVGKASATASGDWNTTIAcfVAILIGLCLTLLLLIAfKKALPALPISITfGLVfyFATDYLvQP
FMDQLAFHQFYI*

SEQUENCE ID NO. 134

FILE NAME: MARMP.UPD

ACCANACANCGGCAGCTGAGGCGGAAACCTAGGCTGCGAGCCGGCCGCGGGCGCGGAGAGAGA
AGGAACCAACACAAGACAGCAGCCCTTCGAGGTCTTTAGGCAGCTTGGAGGAGAAACACATGAGAG
AAAGAATCCCAAGAGGTTTTGTTTTCTTTGAGAAGGTATTTCTGTCCAGCTGCTCCAATGACAGA
GATACCTGCACCTTTGTCTACTTCCAGAATGCCAGATGTCTGAGGACAGCCACTCCAGCAGCG
CCATCCGGAGCCAGAATGACAGCCAAGAACGGCAGCAGCATGACAGGCAGAGACTTGACAAC
CCTGAGCCAATATCTAATGGGCGGCCCCAGAGTAACTCAAGACAGGTGGTGGAAACAAGATGAGGA
GGAAGACGAAGAGCTGACATTGAAATATGGAGCCAAGCATGTCATCATGCTCTTTGTCCCCGTGA
CCCTCTGCATGGTCGTCTGTCTGCGCCACCATCAAAATCAGTCAGCTTCTATACCCGGAAGGACGGT
CAGCTAATCTACACCCCATTCACAGAAGACACTGAGACTGTAGGCCAAAGAGCCCTGCACTCGAT
CCTGAATGCGGCCATCATGATCAGTGTCTATTGTCTATTATGACCATCCTCCTGGTGGTCTGTATA
AATACAGGTGCTACAAGGTCTACAGCCTGGCTTATTATTTCTCTGTGTTGCTGTTCTTT
TTTTCTGTTCAATTAAGTTAGGGGAAGTATTTAAGACCTACAATGTGCGCGTGGACTACGTTACAGT
AGCACTCCTAATCTGGAATTTTGGTGTGGTGGGATGATTGCCATCCACTGGAAAGGCCCCCTTC
GACTGCAGCAGGCGTATCTCATTATGATCAGTGCCCTCATGGCCCTGGTATTTATCAAGTACCTC
CCCGAATGGACCGCATGGCTCATCTTGGCTGTGATTTCAAGTATATGATTGGTGGCTGTTTTATG
TCCCAAAGGCCCCACTTCGTATGCTGGTTGAAACAGCTCAGGAAAGAAATGAGACTCTCTTCCAG
CTCTTATCTATTCTCAACAATGGTGTGGTTGGTGAATATGGCTGAAGGAGACCCAGAAGCCCCAA
AGGAGGGTACCCAAAGAACCCCAAGTATAACACACAAAGAGCGGAGAGAGAGACACAGGACAGTGG
TTCTGGGAACGATGATGGTGGCTTCAGTGAGGAGTGGGAGGCCCCAAAGAGACAGTCACCTGGGGC
CTCATCGCTCCACTCCCGAGTCAAGAGCTGCTGTCCAGGAACTTTCTGGGAGCATTCTAACGAGT
GAAGACCCGGAGGAAAGAGGAGTAAACTTGGACTGGGAGATTTCAATTTCTACAGTGTCTGGT
TGGTAAGGCCTCAGCAACCGCCAGTGGAGACTGGAACACAACCATAGCCTGCTTTGTAGCCATAC
TGATCGGCCTGTGCCTTACATTACTCCTGCTCGCCATTTTCAAGAAAGCGTTGCCAGCCCTCCCC
ATCTCCATCACCTTCGGGCTCGTGTCTACTTCGCCACGGATTACCTTGTGCAGCCCTTCATGGA
CCAACTTGCAATCCATCAGTTTTATATCTAGCCTTTCTGCAGTTAGAACATGGATGTTTCTTCTT
TGATTATCAAAAAACAACAAACAGAGAGCAAGCCGAGGAGGAGACTGGTGACTTTCCTGTGTCC
TCAGCTAACAAAGGCAGGACTCCAGCTGGACTTCTGCAGCTTCCTTCCGAGTCTCCCTAGCCACC
CGCACTACTGGACTGTGGAAGGAAGCGTCTACAGAGGAACGGTTTCCAACATCCATCGCTGCAGC
AGACGGTGTCCCTCAGTGACTTGAGAGACAAGGACAAGGAAATGTGCTGGGCCAAGGAGCTGCCG
TGCTCTGCTAGCTTTGACCGTGGGCATGGAGATTTACCCGCACTGTGAACCTCTCTAAGGTAAACA
AAGTGAGGTGAACC

new

SEQUENCE ID NO. 135

FILE NAME: MARMP.PRO

MTEIPAPLSYFQNAQMSSEDSHSSSAIRSONDSQERQQQHDRQRLDNPEPISNGRPQSNSRQVVEQ
DEEEDDEELTKYGAKHVIMLFVPVTLCMVVVVATIKSVSFYTRKDGQLIYTPFTEDTETVGQAL
HSILNAAIMISVIVIMTILLVLYKYRCYKVIHAWLISSLLLLFFFSFIYLGVEVFKTYNVAVDY
VTVALLIWNFGVVGMIAIHWKGPLRLQAYLIMISALMALVFIKYLPEWTAWLILAVISVYDLVA
VLCPKGPIRLMIVETAQERNETLFPALIYSSTMVWLVNMAEGDPEAQRVPKPKYNTQRAERETQ
DSGSGNDDGGFSEWEAQRDShLGPHRSTPESRAAVQELSGSILTSEDPEERGVLGLGDFIFYS
VLVGKASATASGDWNTTIACFVAILIGLCLTLLLLAIFKKALPALPISITFGLVIFYFATDYLVP
FMDQLAFHQFYI*

SEQUENCE ID NO:136

10	20	30	40	50	60
GAATTCGGCA	CGAGGGCATT	TCCAGCAGTG	AGGAGACAGC	CAGAAGCAAG	CTTTTGGAGC
70	80	90	100	110	120
TGAAGGAACC	TGAGACAGAA	GCTAGTCCCC	CCTCTGAATT	TTACTGATGA	AGAAACTGAG
130	140	150	160	170	180
GCCACAGAGC	TAAAGTGACT	TTTCCCAAGG	TCGCCCAGCG	AGGACGTGGG	ACTTCTCAGA
190	200	210	220	230	240
CGTCAGGAGA	GTGATGTGAG	GGAGCTGTGT	GACCATAGAA	AGTGACGTGT	TAAAAACCAG
250	260	270	280	290	300
CGCTGCECTC	TTTGAAAGCC	AGGGAGCATC	ATTCAATTAG	CCTGCTGAGA	AGAAGAAACC
310	320	330	340	350	360
AAGTGTCCGG	GATTCAAGAC	CTCTCTGCGG	CCCCAAGTGT	TCGTGGTGCT	TCCAGAGGCA
370	380	390	400	410	420
GGGCTATGCT	CACATTTCATG	GCCTCTGACA	GCGAGGAAGA	AGTGTGTGAT	GAGCGGACGT
430	440	450	460	470	480
CCCTAATGTC	GGCCGAGAGC	CCCACGCCGC	GCTCCTGCCA	GGAGGGCAGG	CAGGGCCCAG
490	500	510	520	530	540
AGGATGGAGA	GAATACTGCC	CAGTGGAGAA	GCCAGGAGAA	CGAGGAGGAC	GGTGAGGAGG
550	560	570	580	590	600
ACCCCTGACCG	CTATGTCTGT	AGTGGGGTTC	CCGGGCGGCC	GCCAGGCCTG	GAGGAAGAGC
610	620	630	640	650	660
TGACCCCTCAA	ATACGGAGCG	AAGCATGTGA	TCATGCTGTT	TGTGCCCTGTC	ACTCTGTGCA
670	680	690	700	710	720
TGATCGTGGT	GGTAGCCACC	ATCAAGTCTG	TGCGCTTCTA	CACAGAGAAG	AATGGACAGC
730	740	750	760	770	780
TCATCTACAC	GCCATTCACT	GAGGACACAC	CCTCGGTGGG	CCAGCGCCTC	CTCAACTCCG
790	800	810	820	830	840
TGCTGAACAC	CCTCATCATG	ATCAGCGTCA	TCGTGGTTAT	GACCATCTTC	TTGGTGGTGC
850	860	870	880	890	900
TCTACAAGTA	CCGCTGCTAC	AAGTTCATCC	ATGGCTGGTT	GATCATGTCT	TCACTGATGC
910	920	930	940	950	960
TGCTGPTCCT	CTTCACCTAT	ATCTACCTTG	GGGAAGTGCT	CAAGACCTAC	AATGTGGCCA
970	980	990	1000	1010	1020
TGGACTACCC	CACCCTCTTG	CTGACTGTCT	GGAAGTTCGG	GGCAGTGGGC	ATGGTGTGCA
1030	1040	1050	1060	1070	1080
TCCACTGGAA	GGGCCCTCTG	GTGCTGCAGC	AGGCCTACCT	CATCATGATC	AGTGCCTCA
1090	1100	1110	1120	1130	1140
TGGCCCTAGT	GTTTCATCAAG	TACCTCCAG	AGTGGTCCGC	GTGGGTGATC	CTGGGCGCCA
1150	1160	1170	1180	1190	1200
TCTCTGTGTA	TGATCTCGTG	GCTGTGCTGT	GTCCCAAAGG	GCCTCTGAGA	ATGCTGGTAG
1210	1220	1230	1240	1250	1260
AAACTGCCCA	GGAGAGAAAT	GAGCCCATAT	TCCCTGCCCT	GATATACTCA	TCTGCCATGG

1270	1280	1290	1300	1310	1320
TGTGGACGGT	TGGCATGGCG	AAGCTGGACC	CCTCCTCTCA	GGGTGCCCTC	CAGCTCCCCT
1330	1340	1350	1360	1370	1380
ACGACCCGGA	GATGGAAGAA	GACTCCTATG	ACAGTTTTTG	GGAGCCTTCA	TACCCCGAAG
1390	1400	1410	1420	1430	1440
TCTTTGAGCC	TCCCTTGACT	GGCTACCCAG	GGGAGGAGCT	GGAGGAAGAG	GAGGAAAGGG
1450	1460	1470	1480	1490	1500
GCGTGAAGCT	TGGCCTCGGG	GACTTCATCT	TCTACAGTGT	GCTGGTGGGC	AAGGCGGCTG
1510	1520	1530	1540	1550	1560
CCACGGGCAG	CGGGGACTGG	AATACCACGC	TGGCCTGCTT	CGTGGCCATC	CTCATTGGCT
1570	1580	1590	1600	1610	1620
TGTGTCTGAC	CCTCCTGCTG	CTTGCTGTGT	TCAAGAAGGC	GCTGCCCGCC	CTCCCCATCT
1630	1640	1650	1660	1670	1680
CCATCACGTT	CGGGCTCATC	TTTTACTTCT	CCACGGACAA	CCTGGTGCCG	CCGTTTCATG
1690	1700	1710	1720	1730	1740
ACACCCTGGC	CTCCCATCAG	CTCTACATCT	GAGGGACATG	GTGTGCCACA	GGCTGCAAGC
1750	1760	1770	1780	1790	1800
TGCAGGGAAT	TTTCATTGGA	TGCAGTTGTA	TAGTTTTACA	CTCTAGTGCC	ATATATTTTT
1810	1820	1830	1840	1850	1860
AAGACTTTTC	TTTCCTTAAA	AAATAAAGTA	CGTGTTTACT	TGGTGAGGAG	GAGGCAGAAC
1870	1880	1890	1900	1910	1920
CAGCTCTTTG	GTGCCAGCTG	TTTCATCACC	AGACTTTGOC	TCCCGCTTTG	GGGAGCGCCT
1930	1940	1950	1960	1970	1980
CGCTTCACGG	ACAGGAAGCA	CAGCAGTTTT	ATCCAGATGA	ACTGAGAAGG	TCAGATTAGG
1990	2000	2010	2020	2030	2040
GTGGGGAGAA	GAGCATCCGG	CATGAGGGCT	GAGATGCCCA	AAGAGTGTGC	TCGGGAGTGG
2050	2060	2070	2080	2090	2100
CCCCCTGGAC	CTGGGTGCTC	TGGCTGGAGA	GGAAAAGCCA	GTTCCCTACG	AGGAGTGTTC
2110	2120	2130	2140	2150	2160
CCAATGCTTT	GTCCATGATG	TCCTTGTTAT	TTTATTTCCY	TTANAACTG	ANTCCTNTTN
2170	2180	2190	2200	2210	2220
TTNTTDCGGC	AGTCACMCTN	CTGGGRAGTG	GCTTAATAGT	AANATCAATA	AANAGNTGAG
2230	2240	2250	2260	2270	2280
TCCTNTTAGA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA
2290	2300	2310	2320	2330	2340
AAAAA.....

SEQUENCE ID NO:137

10	20	30	40	50	60
MLTFMADSE	KEVCDERTSL	MSAESPTFRS	COBGRQPED	GENTAQRSSQ	ENERDGEEDP
70	80	90	100	110	120
DRYVCSGVFG	RPPGLRELT	LKYGAHVIM	LFVFTLCMI	VVATIKSVR	FYTERNGQLI
130	140	150	160	170	180
YTPFTEDTPS	VQQRLLNSVL	NULIMISVIV	VMTFIVLVLY	KYRCYKFIHG	NLIMSSILMLL
190	200	210	220	230	240
FLFTYTYLGE	VLKTYNVAMD	YPTLLATVMN	FGAVGVCIH	WKGFLVLQQA	YLMISALMA
250	260	270	280	290	300
LVFIKYLPSW	SAWVILGAIS	VYDLVAVLCP	KGFLEMLVET	AQERNEPIFP	ALIYSSAMVV
310	320	330	340	350	360
TVGMAKLDPS	SQCALQLPYD	PEMEKDSYDS	FGPSPYPEVF	EPPLATGYPCR	ELEERERKGV
370	380	390	400	410	420
KLGLEDPIFY	SVLVGKAAT	GSGDWNTTIA	CFVALILGIC	LTLILLAVFK	KALPALPISI
430	440	450	460	470	480
TFGLIPVFT	INLVREPMDT	LASHQLYI*

Primers

969	ggtaccgccaccatgacagaggtacctgcac	Sequence ID No: 138
970	gaattcactggctgtagaaaaagac	Sequence ID No: 139
989	ggatccgggtccacttcgtatgctg	Sequence ID No: 140
990	ttttttgaattcttaggctatggttgtgttcca	Sequence ID No: 141
994	gattagtgggtgttttgtg	Sequence ID No: 142
995	gattagtggctgttttgtg	Sequence ID No: 143
1003	ttttccagctctcattta	Sequence ID No: 144
1004	ttttccagttctcattta	Sequence ID No: 145
999	tacagtgttctgggttgga	Sequence ID No: 146
996	aaacttggattgggagat	Sequence ID No: 147
100	tacagtgttgtgggttgga	Sequence ID No: 148

SEQUENCE ID NO. 149

FILE NO. 874-984.GEN

GTCTAGATAAGNCAACATTTCAGGGGTAGAAGGGGACTGTTTATTTTTTCCTTTAGTCTCTCTTA
AAGAGTGAGAAAAATTTTCCCAGGAATCCCGGTGGACTTTGCTTCACCACTCATAGGTTCTATACC
AAGTTACAACCCCAACCTTAGAGCTTTTGTAGGAAGAGGCTTGGTGGGATACCGTGCTTGG
CTTGGCTTGGTCAGGATTCACCACCAGAGTCATGTGGGAGGGGGTGGGAACCCAAACAATTCAGG
ATTCTGCCCTCAGGAAATAAAGGAGAAAAATAGCTGTTGGATAAACTACCAGCAGGCACTGCTACA
GCCCATGCTTTGTGGTTTAAGGGCCAGCTAGTTACAATGACAGCTAGTTACTGTTTCCATGTAAT
TTTCTTAAAGGTATTAAATTTTTCTAAATATTAGAGCTGTAACCTCCACTTTCTCTTGAAGGCAC
AGWAAGGGAGTCACAAGACACTGTTGCAGAGAATGATGATGGCGGGTTCAGTGAGGAATGGGAAS
CCCAGRGGGACANTCATCTAGGGCTCATCGCTCTACACCTGAGTCACGAGCTKCTNTCCAGGRA
CTTTCCANCAGTATCCTCGCTGGTGAAGACCCAGAGGAAAGNATGTTTCANTTCTCCATNTTTCAA
AGTCATGGATTCCTTTAGGTAGCTACATTATCAACCTTTTTGAGAATAAAATGAATTGAGAGTGT
TACAGTCTAATTCTATATCACATGTAACCTTTTATTTGGATATATCAGTAATAGTGCTTTTTYNTT
TTTTTTTTTTTTTTTTTTTTTTTTTTNGNGANAGAGTCTCGCTCTGTGCGCCAGGTTGGAGTGCAA
TGGTGCGATCTTGGCTCACTGAAAGCTCCACCNCCCGGTTCAAGTGATTCTCCTGCCTCAGCCN
CCCAAGTAGNTGGGACTACAGGGGTGCGCCACCACGCTGGGATAATTTTGGGNTTTTAGTAGA
GATGGCGTTTCACCANCTTGNGCAGGCTGGTCTTGGAACCTCTGANATCATGATCTGCCTGCCT
TAGCCTCCCCAAAGTGCTGGGATTNCAGGGGTGAGCCACTGTTCTGGGCCTC

SEQUENCE ID NO. 150

FILE NO. 885-1012.GEN

CTGCAGTGAGCCGAGATCATGCTGCTGTACTCCAGCCTGGGCCACAGAGCCAACTCCATCTCCC
AAAAAAAAAAAAATATTAATTAATATGATNAAATGATGCCTATCTCAGAACTCTTGTAAGGATTTT
TTAGKACAAGTGCTGGGTATAAACTATANATTCTATAGATGNCGATTATTACTTAYTATTGTTAT
TGATAAATAACAGCAGCATCTACAGTTAAGACTCCAGAGTCAGTCACATAGAATCTGGNACTCCT
ATTGTAGNAAACCCCNMMAGAAAAGAAACACAGCTGAAGCCTAATTTTGTATATCATTACTGAC
TTCTCTCATTCAATTGTGGGGTTGAGTAGGGCAGTGATATTTTGAATTGTGAAATCATANCAAAG
AGTGACCAACTTTTTAATATTTGTAACCTTTCTTTTTAGGGGGAGTAAACTTGGATTGGGAGA
TTTCATTTTCTACAGTGTTCTGGTTGGTAAAGCCTCAGCAACAGCCAGTGGAGACTGGAACACAA
CCATAGCCTGTTTCGTAGCCATATTAATTGTMSTATACTAATAAGAATGTGTGAGAGCTCTT
AATGTCMAAACTTTGATTACACAGTCCCTTTAAGGCAGTTCGTTTTTAACCCAGGTGGGTAAA
TATTCAGCTATGTGAGGAGCTTTTNGATAATTGGACCTCACCTTAGTAGTTCTCTACCCTGGCC
ACACATTAGAATCACTTGGGAGCTTTTAAACTGTAAAGCTCTGCCCTGAGATATTCTTACTCAAT
TTAATTGTGTAGTTTTTAAATTTCCCGAGGAAATTCTGGTATTTCTGTTTAGGAACCGCTGCCTC
AAGCCTAGCAGCACAGATATGTAGGAAATTAGCTCTGTAAGGTTGGTCTTACAGGGATAAACAGA
TCCTTCCTTAGTCCCTGGACTTAATCACTGAGAGTTTGGGTGGTGGTTTTGGATTTAATGACACA
ACCTGTAGCATGCAGTGTTACTTAAGAC

SEQUENCE ID NO. 151

FILE NO. 901-912.GEN

GGATCCCTCCCCTTTTTAGACCATAACAAGGTAACCTCCGGACGTTGCCATGGCATCTGTAAACTG
TCATGGTGTGGCGGGGAGTGTCTTTTAGCATGCTAATGTATTATAATTAGCGTATAGTGACGAG
TGAGGATAACCAGAGGTCACCTCTCCTCACCATCTTGGTTTTGGTGGGTTTTGGCCAGCTTCTTTA
TTGCAACCAGTTTTATCAGCAAGATCTTTATGAGCTGTATCTTGTGCTGACTTCCTATCTCATCC
CGNAACTAAGAGTACCTAACCTCCTGCAAATTGMAGNCCAGNAGGTCTTGGNCTTATTTNACCCA
GCCCCATTATTCAARATAGAGTNGYTCTTGGNCCAAACGCCYCTGACACAAGGATTTTAAAGTCTTA
TTAATTAAGGTAAGATAGKTCCTTGSATATGTGGTCTGAAATCACAGAAAGCTGAATTTGAAAAA
AGGTGCTTGGASCTGCAGCCAGTAAACAAGTTTTCATGCAGGTGTCAGTATTTAAGGTACATCTC
AAAGGATAAGTACAATTGTGTATGTTGGGATGAACAGAGAGAATGGAGCAANCCAAGACCCAGGT
AAAAGAGAGGACCTGAATGCCTTCAGTGAACAATGATAGATAATCTAGACTTTTAAACTGCATAC
TTCCTGTACATTGTTTTTTCTTGCTTCAGGTTTTTAGAACTCATAGTGACGGGTCTGTTGTTAAT
CCCAGGTCTAACCGTTACCTTGATTCTGCTGAGAATCTGATTTACTGAAAATGTTTTTCTTGTC
TTATAGAATGACAATAGAGAACGGCAGGAGCACAACGACAGACGGAGCCTTGGCCACCCTGANCC
ATTATCTAATGGACGACCCAGGGTAACTCCCGGCAGGTGGTGGANCAAGATGAGGAAGAAGATGA
GGANCTGACATTGAAATATGNCGSCAAGCATGTGATCATGCTCTTGKCCCTGTGACTCTCTGCA
TGGTGGTGGTCGTGGNTACCATTAAAGTCAGTCAGCTTTTATACCCGGAAGGATGGGCAGCTGTAC
GTATGAGTTTKGTTTTATTATTCTCAAASCCAGTGTGGCTTTTCTTTACAGCATGTATCATCAC
CTTGAAGGCCTCTNCATTGAAGGGGCATGACTTAGCTGGAGAGCCCATCCTCTGTGATGGTCAGG
AGCAGTTGAGAGANCGAGGGGTATTACTTCATGTTTTAAGTGGAGAAAAGGAACACTGCAGAAG
TATGTTTTCTGTATGGTATTACTGGATAGGGCTGAAGTTATGCTGAATTGAACACATAAATTCTT
TTCCACCTCAGGGNATTGGGCGCCCATGNTCTTCTGCCTAGAATATTCTTCTTTNCTNACT
TKGGNGGATTAAATTCTGTATCCCCCTCCTCTTGGTGTATATATAAAGTNTTGGTGCCGCAA
AAGAAGTAGCACTCGAATATAAAATTTTCTTTTAATTCTCAGCAAGGNAAGTTACTTCTATATA
GAAGGGTGCACCCNTACAGATGGAACAATGGCAAGCGCACATTTGGGACAAGGGAGGGGAAAGGG
TTCTTATCCCTGACACACGTGGTCCCGCTGNTGTGTNCTNCCCCCACTGANTAGGGTTAGACTG
GACAGGCTTAAACTAATTCCAATTGGNTAATTTAAAGAGAATNATGGGGTGAATGCTTTGGGAGG
AGTCAAGGAAGAGNAGGTAGNAGGTAACCTGAATGA

SEQUENCE ID NO. 152

FILE NO. 910-915.GEN

CNCGTATAAAAGACCAACATTGCCANCNACAACCACAGGCAAGATCTTCTCCTACCTTCCCCCN
GGTGTAATACCAAGTATTCNCCAATTTGTGATAAACTTTTCATTGGAAAGTGACCACCCTCCTTGG
TTAATACATTGTCTGTGCCTGCTTTCACACTACAGTAGCACAGTTGAGTGTTTGCCCTGGAGACC
ATATGACCCATAGAGCTTAAAATATTTCAGTCTGGCTTTTTACAGAGATGTTTCTGACTTTGTAA
TAGAAAATCAACCCAACTGGTTTAAATAATGCACATACTTTCTCTCTCATAGAGTAGTGCAGAGG
TAGNCAGTCCAGATTAGTASGGTGGCTTCACGTTTCATCCAAGGACTCAATCTCCTTCTTTCTTCT
TTAGCTTCTAACCTCTAGCTTACTTCAGGGTCCAGGCTGGAGCCCTASCCTTCATTTCTGACAGT
AGGAAGGAGTAGGGGAGAAAAGAACATAGGACATGTCAGCAGAATTCTCTCCTTAGAAGTTCCAT
ACACAACACATCTCCCTAGAAGTCATTGCCCTTACTTGTTTCTCATAGCCATCCTAAATATAAGGG
AGTCAGAAGTAAAGTCTKXNTGGCTGGGAATATTGGCACCTGGAATAAAAAATGTTTTTCTGTGAA
TGAGAAAACAAGGGGAAGATGGATATGTGACATTATCTTAAGACAACTCCAGTTGCAATTACTCTG
CAGATGAGAGGCACTAATTATAAGCCATATTACCTTTCTTCTGACAACCACTTGTGAGCCNCST
GGTTTTCTGTGGCAGAATCTGGTTCYATAMCAAGTTCCTAATAANCTGTASCCNAAAAAATTTGAT
GAGGTATTATAATTATTTCAATATAAAGCACCCACTAGATGGAGCCAGTGTCTGCTTCACATGTT
AAGTCCTTCTTTCCATATGTTAGACATTTTCTTGAAGCAATTTTAGAGTGTAGCTGTTTTTCTC
AGGTTAAAAATTCTTAGCTAGGATTGGTGAGTTGGGGAAAAGTGACTTATAAGATNCGAATTGAA
TTAAGAAAAAGAAAATTCTGTGTTGGAGGTGCTAATGTGGKTGGTGATCTYCATTAACACTGANC
TAGGGCTTTKGKGTGTTGKTTTATTGTAGAATCTATACCCCATTCANAGAAGATACCGAGACTGTG
GGCCAGAGAGCCCTGCACTCAATTCTGAATGCTGCCATCATGATCAGNGTCATTGTWGTGATGAC
TANNCTCCTGGTGGTTCWGTATAAATACAGGTGCTATAAGGTGAGCATGAGACACAGATCTTTGN
TTTCCACCCTGTTCTTCTTATGGTTGGGTATTCTTGTACAGTAACTTAACTGATCTAGGAAAGA
AAAAATGTTTTGTCTTCTAGAGATAAGTTAATTTTTAGTTTTCTTCTCCTCACTGTGGAACATT
CAAAAAATACAAAAAGGAAGCCAGGTGCATGTGTAATGCCAGGCTCAGAGGCTGAGGCAGGAGGA
TCGCTTGGGCCCAGGAGTTCACAAGCAGCTTGGGCAACGTAGCAAGACCCTGCCTCTATTAAAGA
AAACAAAAACAAATATTGGAAGTATTTTATATGCATGGAATCTATATGTCATGAAAAAATTAGT
GTAATAATATATATATTATGATTAGNTATCAAGATTTAGTGATAATTTATGTTATTTTGGGATTTT
AATGCCTTTTTTAGGCCATTGTCTCAMAAAATAAAGCAGAAAAACAAAAAAGTTGTAAGTAAAA
ATAAACATTTCCATATAATAGCACAATCTAAGTGGGTTTTTGNTTGTGTTTGNNTGTTGAAGC
AGGGCCTTGCCCTNYCACCCAGGNTGGAGTGAAGTGCAGTGGCACGATTTTGGCTCACTGCAG

SEQUENCE ID NO. 153

FILE NO. 917-936.GEN

ATGTTTGACAATTTCTCCGTTCCACCCTTGATTAAATAAGGTAGTATTCAATTTTTTAAGTTTTAG
CTTTTGGATATATGTGTAAGTGTGGTATGCTGTCTAATGAATTAAGACAATTGGTNCTKTCTTTA
CCCMACANCTGGACMAAGAGCAGGCAAGATNCAANAATCAAGTGACCCAGNCAAACCAGACACAT
TTTCTGCTCTCAGCTAGCTTGCCACCTAGAAAGACTGGTTGTCNAAGTTGGAGTCCAAGAATCGC
GGAGGATGTTTTAAATGCAGTTTCTCAGGTTCTCNCCACCCACCAGAAGTTTTGATTCAATTGAGT
GGTGGGAGAGGGCAGAGATATTTGCGATTTTAAACAGCATTCTCTTGATTGTGATGCAGCTGTTTC
SCAAATAGGTACCTTAAAGAAATGACAGGTGTTAAATTTAGGATGGCCATCGCTTGTATGCCGGG
AGAAGCACACGCTGGGCCCAATTTATATAGGGGCTTTTCGTCTCAGCTCGAGCARCCTCAGAACC
CCGACAACCYACGCCAGCKCTCTGGGCGGATTCCRTCAGKTGGGGAAGSCCAGGTGGAGCTCTGG
KTTCTCCCCGCAATCGTTTTCTCCAGGCCGGAGGCCCGCCCCCTTCTCTGGCTCCTCCCCCTCC
TCCGTGGGCCGNCCGCCAACGACGCCAGAGCCGGAAATGACGACAACGGTGAGGGTTCTCGGGCG
GGGCCTCGGACAGGCAGCTCCGGGGTCCNCGNNWTNACATCGGAAACAAAACAGCGGCTGGTCTG
GAAGGAACCTGAKCTACGACCCGCGGCGGCAGCGGGGCGGCGGGGAAGCGTATGTGCGTGATGGG
GAGTCCGGGCAAGCCAGGAAGGCACCGCGGACATGGGCGGCCGCGGGCAGGGNCCGNCCTTTGT
GGCCGCCCCGGGCCGGAAGCCGGTGTCTTAAAGATGACGGGCGGGGCGCGGCCGTTGGGGCTG
GGGAACCCCGTGTGGGAAACCAGGAGGGGCCGCTTTCTCGGGCTTCGGGCGCGGCCGGGTGG
AGAGAGATTCCGGGGAGCCTTGGTCCGAAATGCTGTTTGCTCGAAGACGTCTCAGGGCGCAGGT
GCCTTGGGCCGGGATTAGTAGCCGTCTGAACTGGAGTGGAGTAGGAGAAAGAGGAAGCGTCTTGG
GCTGGGTCTGCTTGAGCAACTGGTGAACTCCGCGCCTCACGCCCCGGGTGTGTCTTGTCCAGG
GGCGACGAGCATTCTGGGCGAAGTCCGCACGCCTCTTGTTCGAGGCGGAAGACGGGGTCTTGATG
CTTTCTCCTTGGTCCGGACTGTCTCGAGGCATGCATGTCCAGTGAATCTTGTGTTTGCTGCTGCT
TCCCTCTCAGATTCTTCTCACCGTTGTGGTCACTCTGCTTTAGGCATATTAATCCATAGTGGAG
GCTGGGATGGGTGAGAGAATTGAGGTGACTTTTCCATAATTAGGTGAGATGTGATTAGAGTYCG
GATCCTNCGGTGGTGGCAGAGGCTTACCAAGAAACACTAACGGGACATGGGAACCAATTGAGGAT
CCAGGGAATAAAGTGTGAAGTTGACTAGGAGGTTTTCACTTTAAGAACATGGCAGAGACATTCTC
AGAAATAAGGAAGTTAGGAAGAAAGACCTGTTTTAGAGAGGAGGGCGAGGAAGTGGTTTTGGAAGT
GTCACTTTGGAAGTGCCAGCAGGTGAAAATGCCCTGTGAACAGGACTGGAGCTGAAAACAGGAAT
CAATTCATAGATTTCCAGTTGATGTTGGAGCAGTGGAGAAGTCTAANCTAAGGAAGGGGAAGAG
GAGGCCAAGCCAAACACTTAGGAACACTTNCNACGAGGGGTGGAAGAAGAGCAAGGAGCCAGCT
GAGGAGAATGAGTGTGGTTGGAGAACCACCACAGCNCAGGGTCCGCCAGANCTGAGGAAGGGGAGG
GAAGCTTATCGAGKAMSGWCRACMKCGAGTTGGCAGGGAT

SEQUENCE ID NO. 154
FILE NO. 930-919.GEN

GTCTTTCCCATCTTCTCCACAGAGTTTGTGCCTTACATTATTACTCCTTGCCATTTTCAAGAAAG
CATTGTCAGCTCTTCCAATCTCCATCACCTTTGGGCTTGTTTTCTACTTTGCCACAGATTATCTT
GTACAGCCTTTTATGGACCAATTAGCATTCCATCAATTTTATATCTAGCATATTTGCGGTTAGAA
TCCCATGGATGTTTTCTTCTTGACTATAACAAAATCTGGGGAGGACAAAGGTGATTTCTGTGTC
CACATCTAACAAATCAAGATCCCCGGCTGGACTTTTGGAGGTTTCCTTCCAAGTCTTCCTGACCAC
CTTGCACTATTGGACTTTTGAAGGAGGTGCCTATAGAAAACGATTTTGAACATACTTCATCGCAG
TGGACTGTGTCCTCGGTGCAGAACTACCAGATTTGAGGGACGAGGTCAAGGAGATATGATAGGC
CCGGAAGTTGCTGTGCCCCATCAGCAGCTTGACGCGTGGTCACAGGACGATTTTCACTGACACTG
CGAACTCTCAGGACTACCGTTACCAAGAGGTAGGTGAAGTGTTTAAACCAAACGGAACCTTC
ATCTTAAACTACAGTTGAAAATCAACCCAATAATTCTGTATTAAGTGAATTCGAACTTTTCAG
GAGGTACTGTGAGGAAGAGCAGGCACCACCAGCAGAATGGGGAATGGAGAGGTGGGCAGGGGTTT
CAGCTTCCCTTTGATTTTTTG

SEQUENCE ID NO. 155
FILE NO. 932-943.GEN

GGATCCGCCCCGCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCTCCTGGCTGA
GTCTGCGATTTCTTGCCAGCTCTACCCAGTTGTGTCATCTTAAGCAAGTCACTGAACTTCTCTGG
ATTCCCTTCTCCTNNWGTAAAAATAAGNATGTTATCTGNCCNNCCTGCCTTGGGCATTGTGATAAG
GATAAGATGACATTATAGAATNTNGCAAAATTTAAAGCGCTAGACAAATGATTTTATGAAAAATAT
AAAGATTAGNTTGAGTTTGGGCCAGCATAGAAAAAGGAATGTTGAGAACATTCCNTTAAGGATTA
CTCAAGCYCCCCTTTTGSTGKNWAATCAGANNGT CATNNAMNTATCNTNTGTGGGYTGAAAATGT
TTGGTTGTCTCAGGCGGTTCTACTTATTGCTAAAGAGTCCTACCTTGAGCTTATAGTAAATTTG
TCAGTTAGTTGAAAGTCGTGACAAATTAATACATTCTGGTTTACAAATTGGTCTTATAAGTATT
TGATTGGTNTAAATGNATTTACTAGGATTTAACTAACAATGGATGACCTGGTGAAATCCTATTTT
AGACCTAATCTGGGAGCCTGCAAGTGACAACAGCCTTTGCGGTCCTTAGACAGCTTGGCCTGGAG
GAGAACACATGAAAGAMMGGTTTGWNTCTGNTTAWTGTAATCTATGRAAGTGTTTTTWTMACAG
TATAATTGTMGTMACAAAGTTCTGTTTTTCTTCCCTTTNCAGAACCTCAAGAGGCTTTGTTTTT
TGTGAAACAGTATTTCTATACAGNTGCTCCAATGACAGAGTNACCTGCACCGTTGTCTACTTCC
AGAATGCACAGATGTCTGAGGACAACCACCTGAGCAATACTGTACGTAGCCAGGTACAGCGTCAG
TYTCTNAAACTGCCTYYGNCAGACTGGATTCACTTATCATCTCCCCTCACCTCTGAGAAATGCTG
AGGGGGSTAGGNAGGGCTTTCTCTACTTNACCACATTTNATAATTATTTTTTGGGTGACCTTCAGC
TGATCGCTGGGAGGGACACAGGGCTTNTTTAACACATAGGGTGTTGGATACAGNCCCTCCCTAAT
TCACATTTTANC

SEQUENCE ID NO. 156
FILE NO. 951-952.GEN

CTGCAGCTTTCCTTTAAACTAGGAAGACTTGTTCCCTATACCCCAGTAACGATACACTGTACACTA
AGCAAATAGCAGTCAAACCCAAATGAAATTTNTACAGATGTTCTGTGTCATTTTATNTTGTTTAT
GTTGTCTCCCCCACCCTTACCTGCCATTTATTTTCATATTCATTCAACGTCTNNNTGT
GTAAAAAGAGACAAAAAACATTAACTTTTTTTCCTTCGTTAATTCCTCCCTACCACCCATTTACA
AGTTTAGCCCATACATTTTATTAGATGTCTTTTATGTTTTTCTTTTNTAGATTTAGTGGCTGTT
TNGTGTCCGAAAGGTCCACTTCGTATGCTGGTTGAAACAGCTCAGGAGAGAAATGAAACGCTTTT
TCCAGCTCTCATTTACTCCTGTAAGTATTTGGAGAATGATATTGAATTAGTAATCAGNGTAGAAT
TTATCCGGAACTTGAAGANATGTNACTATGGCAATTTCANGGNACTTGTCTCATCTTAAATGANA
GNATCCCTGGACTCCTGNAG

SEQUENCE ID NO. 157

FILE NO. 983-1011.GEN

CCCCGTCNATGCATACTTTGTGTGTCCAGTGCTTACCTGGAATCCNGTCTTTCCCAACAGCAACA
ATGGTGTGGTTGGTGAATATGGCAGAAGGAGACCCGGAAGCTCAAAGGAGAGTATCCAAAAATTC
CAAGTATAATGCAGAAAGTAGGTAACTYYNTTAGATAMNATCTTGATTTTNCAGGGTCACTGTT
ATAAGCTAACAGTATAGNAATGTTTTTATCGTCTTTCTNKGGNCATAGACTCCTNKGAGAATCTC
TTGAGAACTATGATAATGCCCAGTAAATACNCAGATAAGTATTTAAGGAGTNCAGATACTCAAAN
CCCAACAATACNGTCAAAGCATCCTAGGTTAAGACAMCNCATTAAATACAGAATACCAGCATG
GAAAGGTTCAAGCTGAGGTTATGATTGGGTTTGGGNNNGTTTTTATAAGTCATGATT
TTAAAAGAAAAATAAACTCTCTCAAACATGTAAAAGTAAGAATCTCCTAAA

SEQUENCE ID NO. 158
FILE NO 925-913.GEN

CAGGAGTGGACTAGGTAAATGNAAGNTGTTTTAAAGAGAGATGNGGNCNCGGACATAGTGGTACA
CANCTGTAATGCTCANCACATKATGGGGAGTACTGAAGGNGGNSGGATCACTTGNGGGTCNGGAAT
NTGAGANCAGCCTGGGCAANATGGCGAAACCTGTCTCTACTAAAAATAGCCANAAWNWAGCCTA
GCGTGGTGGCGCRACGCGTGGTTCACCTACTCAGGAGGCNTAAGCACGAGNANTNCTTGAACC
CAGGAGGCAGAGGNTGTGGTGARCTGAGATCGTGCCACTGCACTCCAGTCTGGGCGACMAAGTGA
GACCTGTCTCCNNNAAGAAAAAATCTGTACTTTTTAAGGGTTGTGGGACCTGTTAATTAT
ATTGAAATGCTTCTYTTCTAGGTCATCCATGCCTGGCTTATTATATCATCTCTATTGTTGCTGCT
CTTTTTTACATTCACTTACTTGGGGTAAGTTGTGAAATTTGGGGTCTGTCTTTCAGAATTAATA
CCTNNGTGCTGTGTAGCTATCATTTAAAGCCATGTACTTTGNTGATGAATTACTCTGAAGTTTTA
ATTGTNTCCACATATAGGTCATACTTGGTATATAAAAGACTAGNCAGTATTACTAATTGAGACAT
TCTTCTGTNGCTCCTNGCTTATAATAAGTAGAACTGAAAGNAACTTAAGACTACAGTTAATTCTA
AGCCTTTGGGGAAGGATTATATAGCCTTCTAGTAGGAAGTCTGTGCNATCAGAATGTTTNTAAA
GAAAGGTTNTCAAGGAATNGTATAAANACCAAAAATAATTGAT

SEQUENCE ID NO. 159
FILE NO. 849-892.GEN

GTTNTCCNAACCAACTTAGGAGNTTGGACCTGGGGAAGACCNACNTGATCTCCGGGAGGNAAAGA
CTNCAGTTGAGCCGTGATTGCACCCACTTTACTCCAAGCCTGGGCAACCAAAATGAGACACTGGC
TCCAAACACAAAAACAAAAACAAAAAAGAGTAAATTAATTTANAGGGAAGNATTAAATAAATAA
TAGCACAGTTGATATAGGTTATGGTAAAATTATAAAGGTGGGANATTAAATATCTAATGTTTGGGA
GCCATCACATTATTCTAAATAATGTTTTGGTGAAATTATTGTACATCTTTTAAATCTGTGTAA
TTTTTTTTCAGGGAAGTGTTTAAACCTATAACGTTGCTGTGGACTACATTACTGTTNCACTCCT
GATCTGGAATTTTGGTGTGGTGGGAATGATTTCCATTCACTGGAAAGGTCCACTTCGACTCCAGC
AGGCATATCTCATTATGATTAGTGCCCTCATGNCCCTGKTGTTTATCAAGTACCTCCCTGAATGG
ACTGNGTGGCTCATCTTGGCTGTGATTTTCAGTATATGGTAAACCCAAGACTGATAATTTGTTTG
TCACAGGAATGCCCCACTGGAGTGTTTTCTTTCCTCATCTCTTTATCTTGATTTAGAGAAAATGG
TAACGTGTACATCCCATAACTCTTCAGTAAATCATTAATTAGCTATAGTAACTTTTTCATTTGAA
GATTTTCGGCTGGGCATGGTAGCTCATGCCTGTAATCTTAGCACTTTGGGAGGCTGAGGCGGGCAG
ATCACCTAAGCCCAGAGTTCAAGACCAGCCTGGGCAACATGGCAAAACCTCGTATCTACAGAAAA
TACAAAAATTAGCCGGGCATGGTGGTGACACCTGTAGTTCCAGCTACTTAGGAGGCTGAGGTGG
GAGGATCGATTGATCCAGGAGGTCAAGNCTGCAG

WE CLAIM:

1. An isolated DNA molecule comprising a nucleotide sequence coding for Alzheimer Related Membrane Protein (ARMP) or a functional fragment or variant of the protein.
2. The DNA molecule of claim 1 wherein the nucleotide sequence codes for the amino acid Sequence ID No:2.
3. The DNA molecule of claim 1 wherein the nucleotide sequence codes for the amino acid Sequence ID No:133.
4. The DNA molecule of claim 1 wherein the nucleotide sequence codes for the amino acid Sequence ID No:134.
5. The DNA molecule of claim 1 wherein the nucleotide sequence codes for human ARMP and is selected from the group consisting of
 - (a) Sequence ID No:1;
 - (b) nucleotides 186 to 2764 of Sequence ID No:1;
 - (c) Sequence ID No:5;
 - (d) Sequence ID No:132; and
 - (e) nucleotides 1 to 1017 and 1117 to 2791 of sequence ID No:1.
6. The DNA molecule of claim 1 wherein the nucleotide sequence codes for mouse ARMP and has nucleotide Sequence ID No:134.
7. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of
 - (a) a deoxyribonucleotide sequence complementary to Sequence ID No:1;

- (b) a ribonucleotide sequence complementary to Sequence ID No:1;
- (c) a ribonucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b);
- (d) a nucleotide sequence of at least 12 consecutive nucleotides capable of hybridising to Sequence ID No:1; and
- (e) a nucleotide sequence capable of hybridising to a nucleotide sequence of (d).

8. The DNA molecule of claim 1 wherein the nucleotide sequence is the ARMP-coding nucleotide sequence of ATCC Deposit No. _____.

9. An isolated DNA molecule comprising a nucleotide sequence selected from the group consisting of

- (a) Sequence ID No:3;
- (b) Sequence ID Nos:6 to 125;
- (c) Sequence ID Nos:126 to 131;
- (d) Sequence ID Nos:138 to 148; and
- (e) Sequence ID Nos:149 to 159.

10. An isolated DNA molecule comprising a nucleotide sequence coding for E5-1 protein or a functional fragment or variant of the protein.

11. The DNA molecule of claim 10 wherein the nucleotide sequence is selected from the group consisting of:

- (a) a nucleotide sequence coding for amino acid Sequence ID No: 137; and
- (b) nucleotide Sequence ID No: 136.

12. The DNA molecule of claim 10 wherein the nucleotide sequence is the E5-1 coding nucleotide sequence of ATCC Deposit No. _____.

13. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of

- (a) a deoxyribonucleotide sequence complementary to Sequence ID No:134;
- (b) a ribonucleotide sequence complementary to Sequence ID No:134;
- (c) a ribonucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b);
- (d) a nucleotide sequence of at least 12 consecutive nucleotides capable of hybridising to Sequence ID No:134; and
- (e) a nucleotide sequence capable of hybridising to a nucleotide sequence of (d).

14. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of

- (a) a deoxyribonucleotide sequence complementary to Sequence ID No:136;
- (b) a ribonucleotide sequence complementary to Sequence ID No:136;
- (c) a ribonucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b);
- (d) a nucleotide sequence of at least 12 consecutive nucleotides capable of hybridising to Sequence ID No:136; and
- (e) a nucleotide sequence capable of hybridising to a nucleotide sequence of (d).

15. An isolated DNA molecule comprising a nucleotide sequence coding for a mutant form of Alzheimer Related Membrane Protein.

16. The DNA molecule of claim 15 wherein the nucleotide sequence has at least one mutation selected from the group consisting of

- i) 685, A→C ii) 737, A→G iii) 986, C→A
- iv) 1105, C→G v) 1478, G→A vi) 1027, C→T
- vii) 1102, C→T and viii) 1422, C→G.

17. The DNA molecule of claim 15 wherein the nucleotide sequence comprises the nucleotide Sequence ID No:1 or 132 having at least one mutation selected from the group consisting of

- i) 685, A→C ii) 737, A→G iii) 986, C→A
- iv) 1105, C→G v) 1478, G→A vi) 1027, C→T
- vii) 1102, C→T and viii) 1422, C→G.

18. A recombinant vector comprising the DNA molecule of any of claims 1 to 17.

19. A host cell transfected with a recombinant vector comprising the DNA molecule of any of claims 1 to 17.

20. Purified Alzheimer Related Membrane Protein or a functional fragment or variant thereof.

21. The protein of claim 20 comprising an amino acid sequence selected from the group consisting of

- (a) Sequence ID No:2;
- (b) Sequence ID NO:4;
- (c) Sequence ID No:133; and
- (d) Sequence ID No:135.

22. Substantially pure mutant Alzheimer Related Membrane Protein.

23. The protein of claim 22 having a mutation selected from the group consisting of

- i) M 146L; ii) H 163R; iii) A 246E;
- iv) L 286V; v) C 410Y; vi) A 260V;
- vii) A 285V; and viii) L 392V.

24. Substantially pure E5-1 protein.

25. The protein of claim 25 comprising the amino acid Sequence ID No:137.

26. An isolated DNA molecule comprising a splice variant of the nucleotide Sequence ID No:1.

27. The DNA molecule of claim 26 comprising nucleotides 1 to 1017 and 1117 to 2791 of Sequence ID No:1.

28. A nucleotide sequence which codes for an antigenic determinant of a protein selected from the group consisting of Sequence ID No:2, Sequence ID No: 4, Sequence ID No: 133, Sequence ID No: 135 and Sequence ID No: 137.

29. A nucleotide sequence of claim 28, wherein said sequence is selected from the group of nucleotide sequences which code for the following protein fragments of Sequence ID No:2 consisting of amino acid residues 27-44, 46-48, 50-60, 66-67, 107-111, 120-121, 125-126, 155-160, 185-189, 214-223, 220-230, 240-245, 267-269, 273-282, 300-370 and 400-420.

30. A polypeptide of at least 6 amino acid residues selected from at least 6 consecutive amino acid residues of Seq. ID No: 2.

31. A polypeptide of at least 6 amino acid residues selected from at least 6 consecutive amino acid residues of Seq. ID No: 13.

32. A polypeptide having antigenic properties and selected from the group consisting of amino acid residues 27-44, 46-48, 50-60, 66-67, 107-111, 120-121, 125-126, 155-160, 185-189, 214-223, 220-230, 240-245, 267-269, 273-282, 300-370 and 400-420 of Sequence ID No: 2.

33. An antibody capable of specific binding to ARMP.

34. An antibody capable of specific binding to an ARMP extracellular domain.

35. An antibody capable of binding to a polypeptide in accordance with claim 29.

36. A bioassay for determining if a subject has a normal or mutant Alzheimer's related membrane protein (ARMP), said bioassay comprising:

- i) providing a biological sample of said subject;
- ii) conducting a biological assay on said sample to detect a normal or mutant gene sequence coding for ARMP, a normal or mutant ARMP amino acid sequence or a normal or defective protein function.

37. A bioassay of claim 36, wherein said bioassay is a DNA or RNA based biological assay.

38. A bioassay of claim 37, wherein said biological assay is selected from the group consisting of probe hybridization, direct DNA sequencing, restriction enzyme analysis, electrophoretic mobility, RNase detection, chemical cleavage, ligase-mediated detection and PCR amplification.

39. A bioassay of claim 38, wherein said biological assay detects at least one mutatic selected from the group consisting of

- i) 685, A→C ii) 737, A→G iii) 986, C→A
- iv) 1105, C→G v) 1478, G→A vi) 1027, C→T
- vii) 1102, C→T and viii) 1422, C→G.

40. A bioassay of claim 36, wherein said bioassay is amino acid based biological assay.

41. A bioassay of claim 40, wherein said biological assay is selected from the group consisting of immunoassay, enzyme site specific assay, electrophoretic mobility of cleaved polypeptides.

42. A bioassay of claim 41, wherein said selected biological assay detects at least one mutation selected from the group consisting of:

- i) M 146L; ii) H 163R; iii) A 246E;
- iv) L 286V; v) C 410Y; vi) A 260V;
- vii) A 285V; and viii) L 392V.

43. A bioassay of claim 36, wherein said biological assay detects normal or defective protein function.

44. A process for recombinantly producing Alzheimer's related membrane protein (ARMP) comprising culturing a host cell of claim 19 under suitable conditions to produce said ARMP by expressing said DNA sequence.

45. A therapeutic composition comprising Alzheimer's related membrane protein and a pharmaceutically acceptable carrier.

46. A recombinant vector for transforming mammalian tissue cells to express a therapeutically effective amount of Alzheimer's related membrane

protein in said cells, said vector being delivered to said cells by a suitable vehicle.

47. A recombinant vector of claim 46, wherein said tissue cells comprise liver, kidney, spleen, bone marrow and neurological cells.

48. A recombinant vector of claim 46, wherein said neurological cells comprise central nervous system cells of neuron, brain and vascular cell type.

49. A recombinant vector of claim 46 wherein said vehicle is selected from the group consisting of vaccinia virus, adenovirus, retrovirus, liposome transport, neurotropic viruses and Herpes simplex.

50. A method of treating a patient deficient in normal Alzheimer's related membrane protein (ARMP) comprising the step of administering to said patient a therapeutically effective amount of said protein targeted at a variety of patient cells which normally express ARMP.

51. A method of claim 50, wherein said variety of patient cells comprise heart, brain, lung, liver, skeletal muscle, kidney, pancreas and neurological cells.

52. An immunotherapy for treating a patient having Alzheimer's Disease due to cellular production of mutant ARMP, said immunotherapy comprising treating said patient with antibodies specific to said mutant ARMP, said antibodies binding mutant ARMP to reduce thereby levels of mutant ARMP in said patient.

53. An immunotherapy of claim 52, wherein said antibodies are developed by said patient's immune system upon administration to said patient of a vaccine comprising said mutant ARMP and a pharmaceutically acceptable carrier.

54. A vaccine composition for invoking an immune response in a human susceptible Alzheimer's Disease, said composition comprising a mutant ARMP and a pharmaceutically acceptable carrier.

55. A method of treating a patient of Alzheimer's Disease comprising administering to said patient a therapeutically effective amount of a ligand or chemical molecule which corrects symptoms associated with expression of mutant ARMP gene.

56. A transgenic animal model for animal symptoms of Alzheimer's disease, said animal model having within its genome a DNA molecule in accordance with claim 1 with at least one mutation which when expressed results in the presence or absence of mutant ARMP in the animal's cells and thereby manifests the symptoms.

57. A transgenic animal model of claim 56, wherein said mutation is at least one nucleotide mutation.

58. A transgenic animal model of claim 56, wherein said polynucleotide sequence is Sequence ID No: 1.

59. A transgenic animal model of claim 56, wherein said polynucleotide sequence is Sequence ID No: 3.

60. A transgenic animal of claim 58, wherein said sequence mutations are selected from the group consisting of:

- i) 685, A→C ii) 737, A→G iii) 986, C→A
- iv) 1105, C→G v) 1478, G→A vi) 1027, C→T
- vii) 1102, C→T and viii) 1422, C→G.

61. A transgenic animal of claim 58, wherein said animal is a rodent.

62. A transgenic animal of claim 58, wherein said rodent is a mouse.
63. A transgenic mouse model for Alzheimer's disease, said mouse having said mammalian polynucleotide of claim 134 with a mutation which when expressed results in the presence or absence of a mutant ARMP in the animal's cells and thereby manifest the syndrome.
64. A transgenic mouse exhibiting a physiological or neurological disorder which can be linked to one more mutations in the mouse ARMP gene or mouse ARMP.
65. A transgenic mouse of claim 64 exhibiting symptoms of cognitive, memory or behavioural disturbances.
66. A transgenic mouse of claim 65 exhibiting tissue cell disorders in heart, brain, lung, liver, skeletal muscle, kidney, pancreas and neurological cells.
67. Use of a transgenic animal of claim 56 for screening proteins, ligands and chemical molecules for efficacy in reversing effects of Alzheimer's Disease.
68. Use of a transgenic mouse of claim 63 for screening proteins, ligands and chemical molecules for efficacy in reversing effects of Alzheimer's Disease.
69. An isolated DNA molecule comprising a nucleotide sequence coding for a mutant form of E5-1 protein.
70. The DNA molecule of claim 69 wherein the nucleotide sequence has at least one mutation selected from the group consisting of:
- i) 787, A→T and
 - ii) 1080, A→G.

71. Substantially pure mutant E5-1 protein.

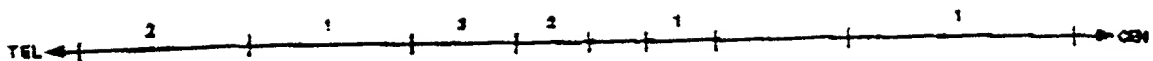
72. The protein of claim 71 having a mutation selected from the group consisting of:

- i) Asn141Ile and
- ii) Met239Val.

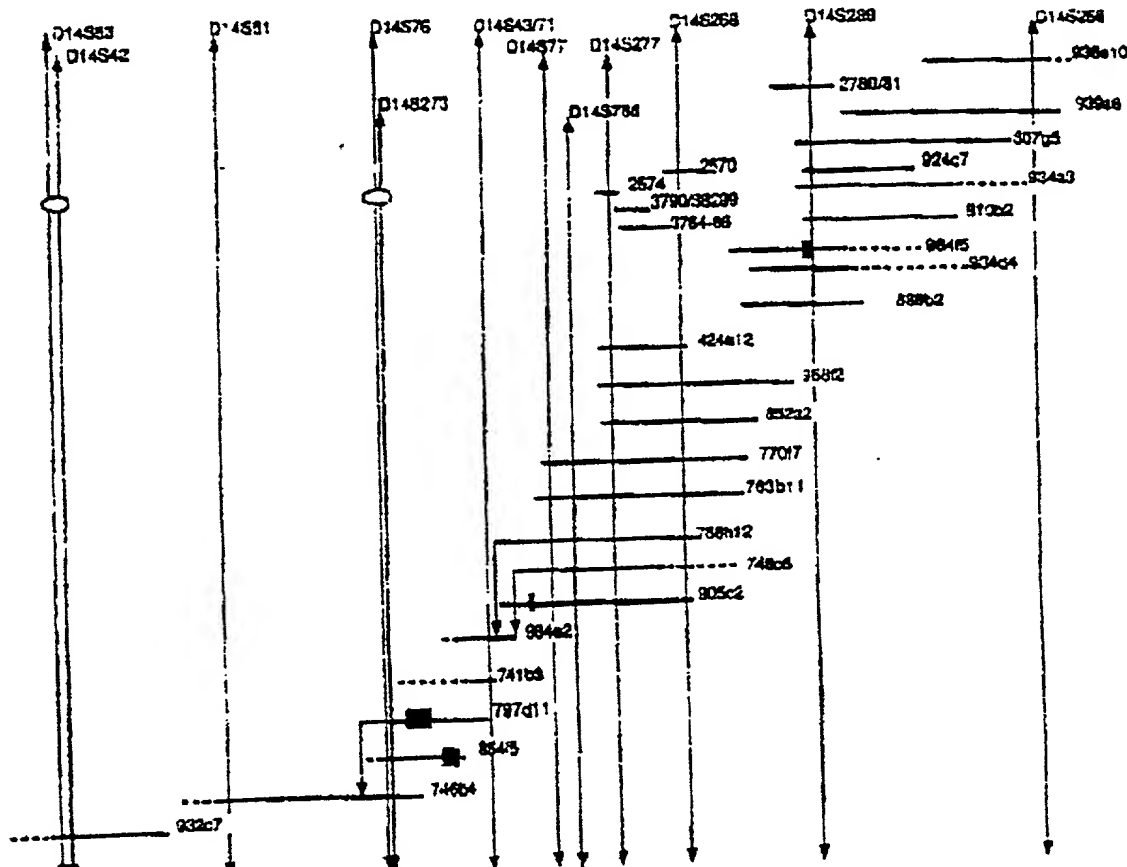
GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE

Abstract of the Disclosure

The present invention describes the identification, isolation, cloning, and determination of the Alzheimer Related Membrane Protein (ARMP) gene on chromosome 14 and a related gene, E5-1, on chromosome 1. Normal and mutant copies of both genes are presented. Transcripts and products of these genes are useful in detecting and diagnosing Alzheimer's disease, developing therapeutics for treatment of Alzheimer's disease, as well as the isolation and manufacture of the protein and the construction of transgenic animals expressing the mutant genes.



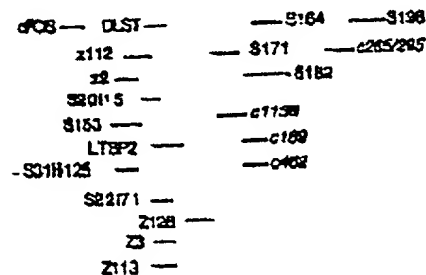
A



Approx. scale: 1 Lt

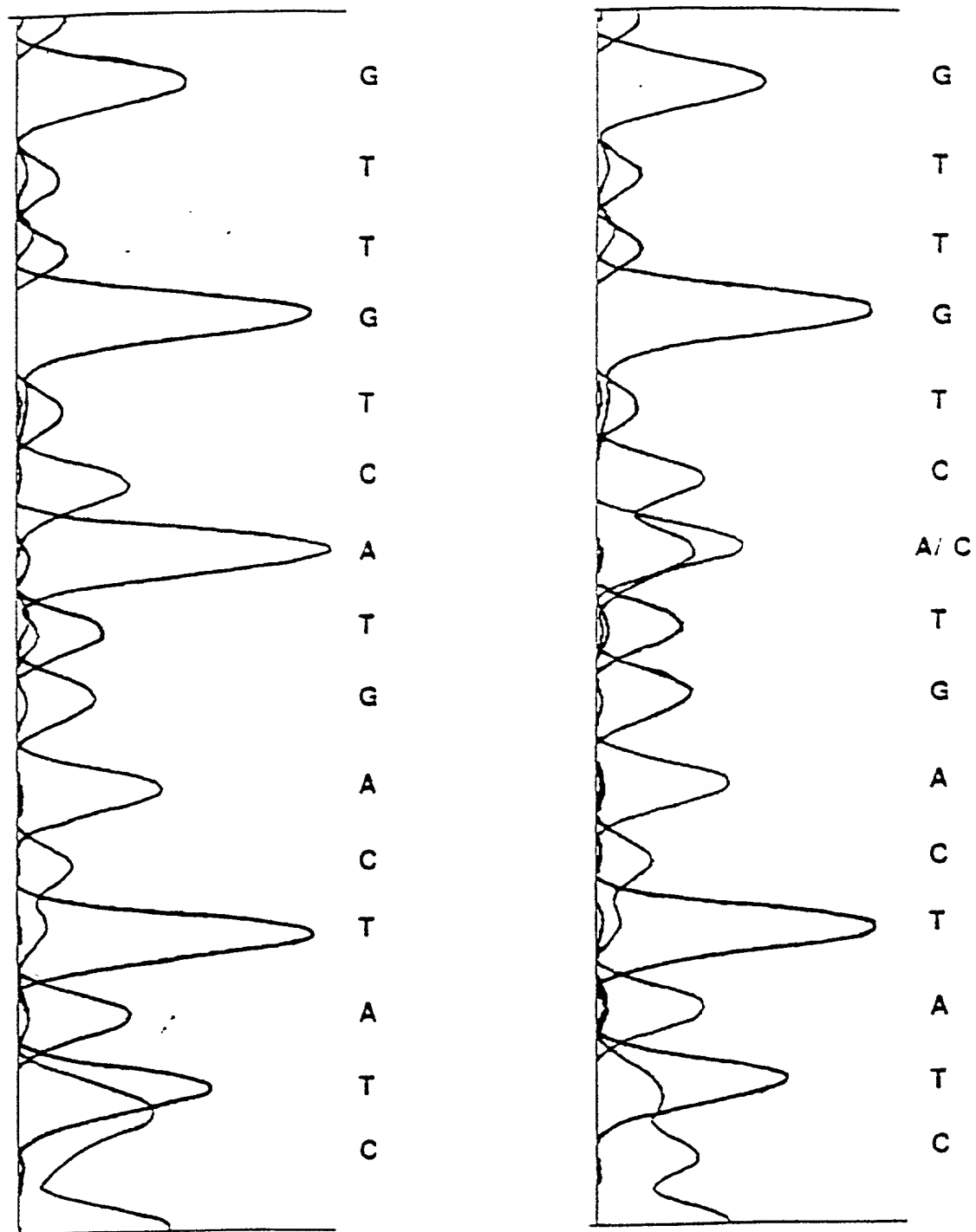


C



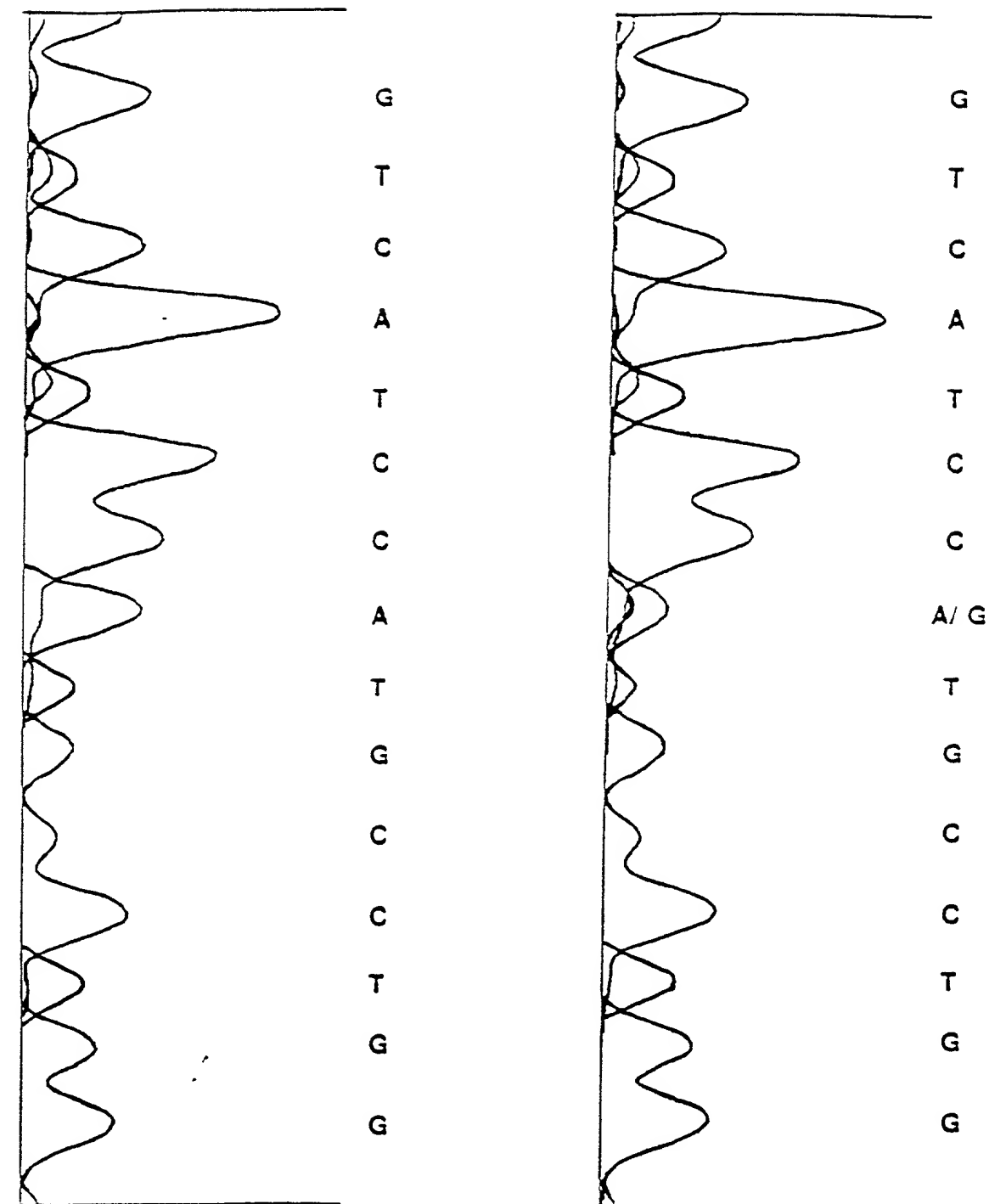
D

FIG. 1



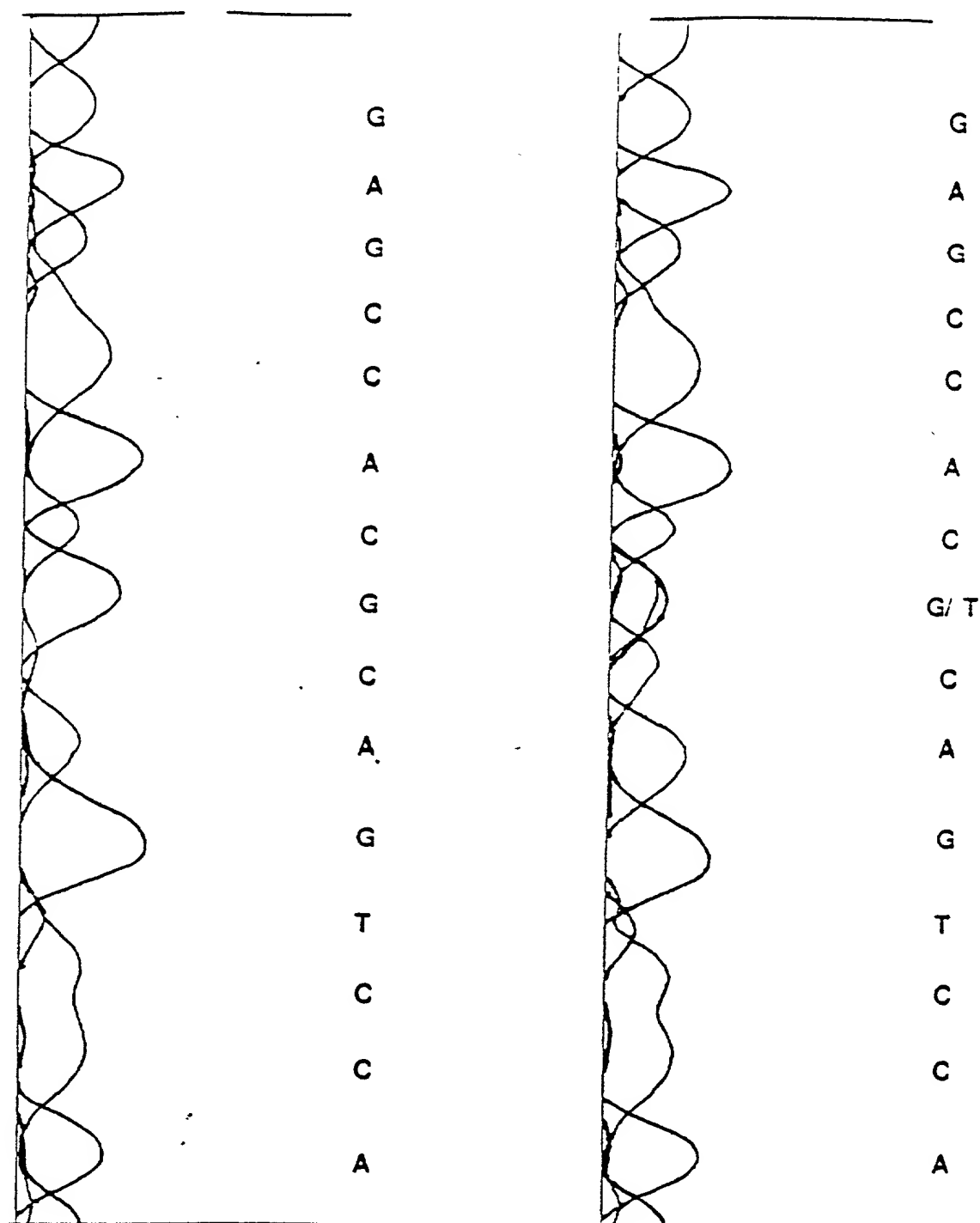
Met 146 Leu

FIG. 2A



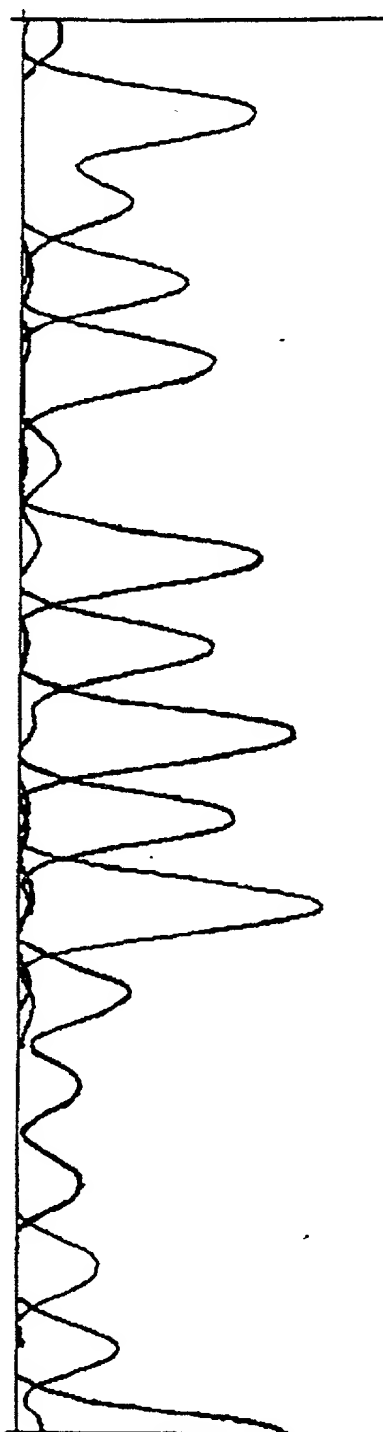
His 163 Arg

FIG. 2B

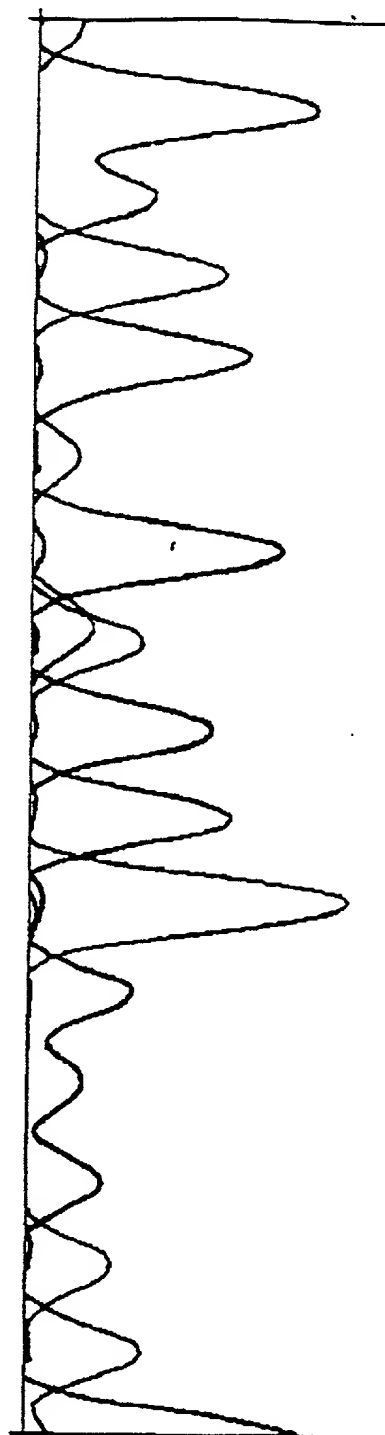


Ala 246 Glu

FIG. 2C



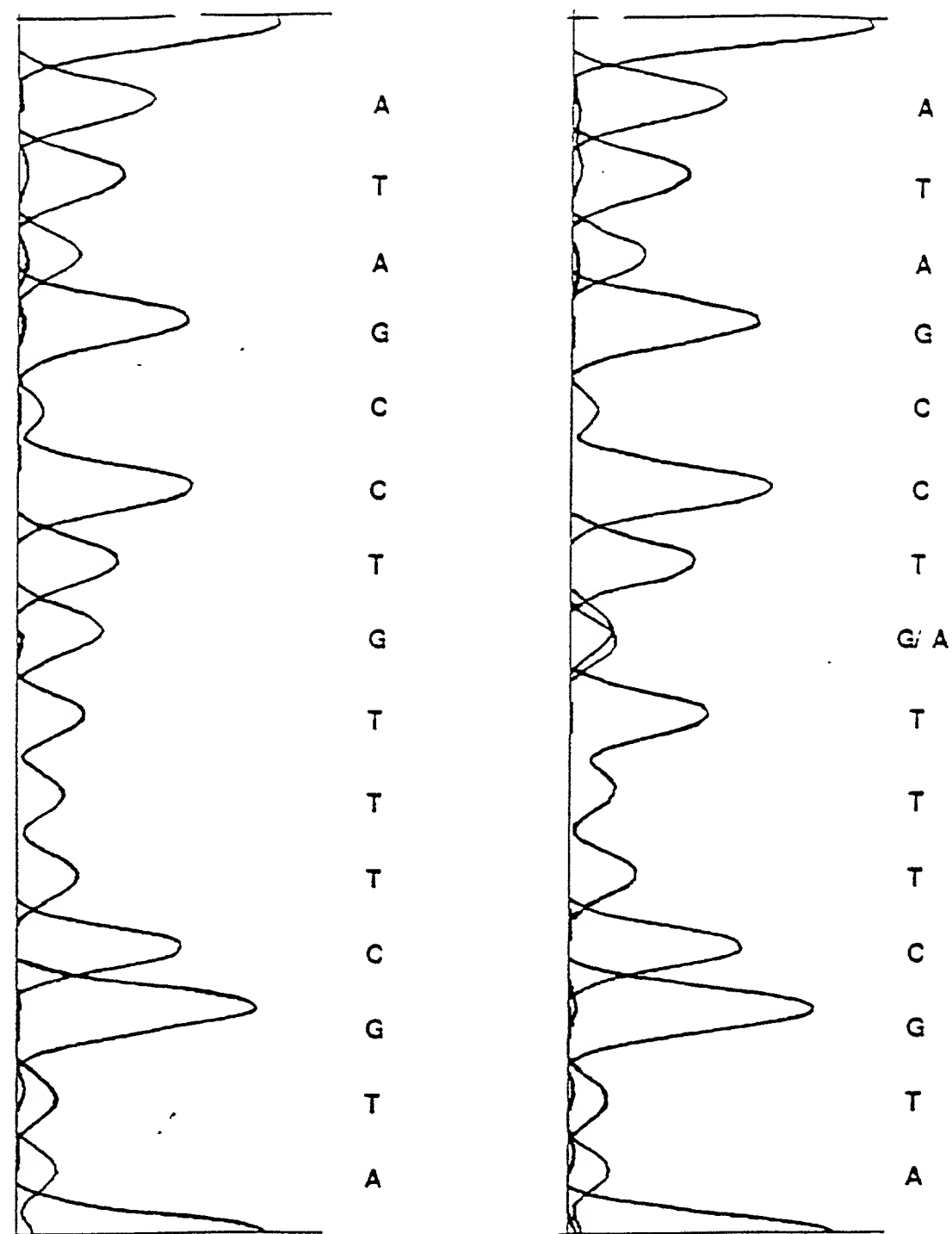
C
C
A
G
C
T
C
T
C
A
T
T
T
A
C



C
C
A
G
C
T
C/G
T
C
A
T
T
T
A
C

Lex 286 Ush

FIG. 2D



Cys 410 Tyr

FIG. 2E

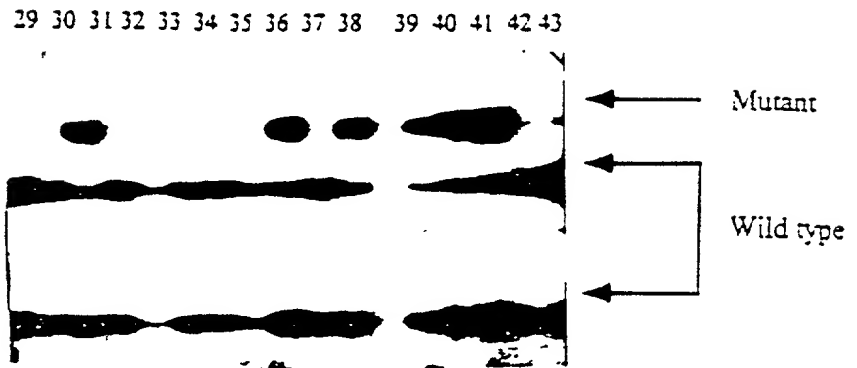
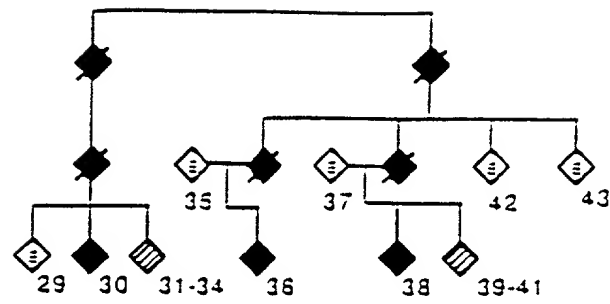


FIG. 3A

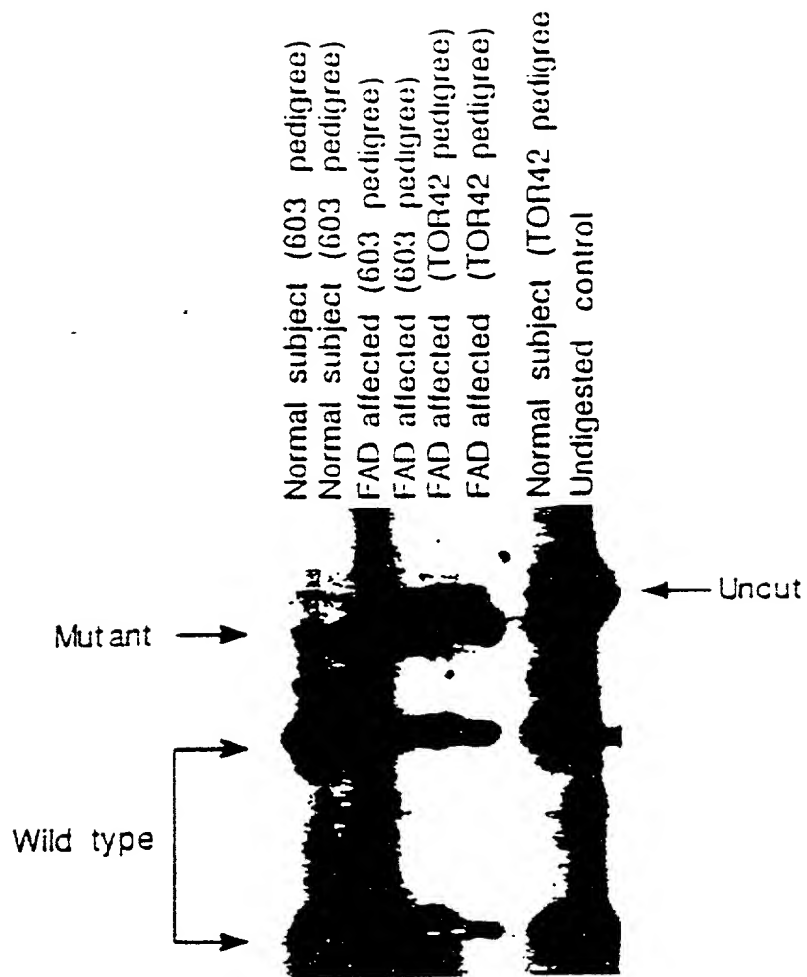
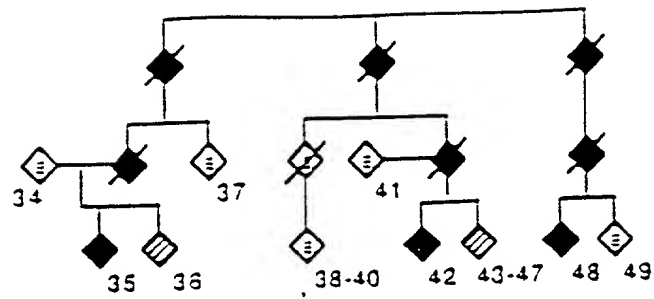


FIG. 3B



34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49

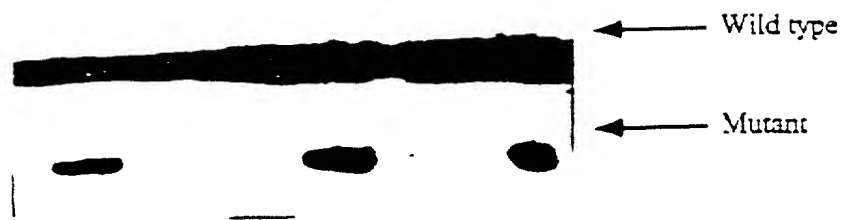


FIG. 3C

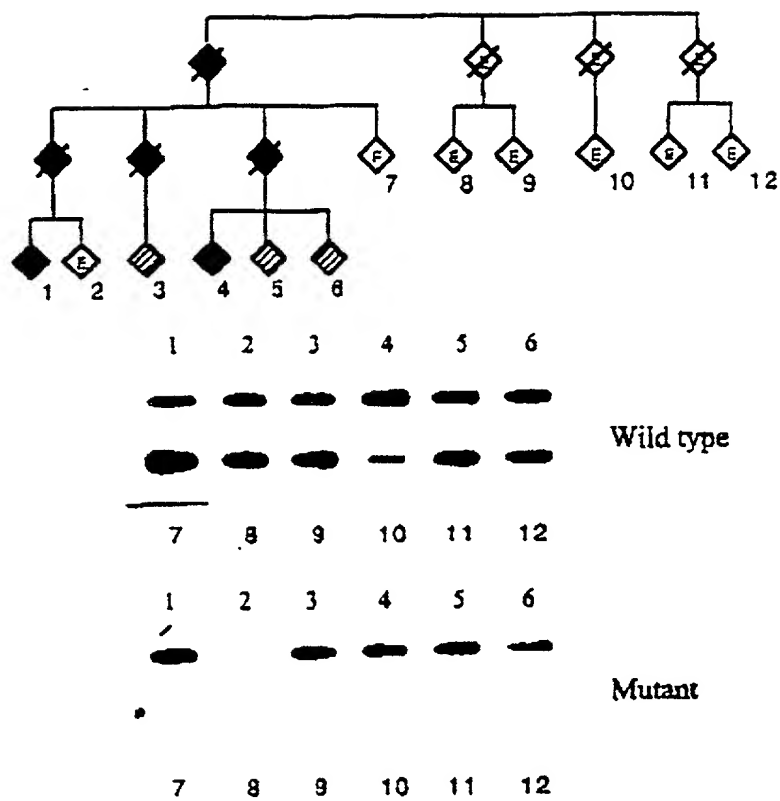
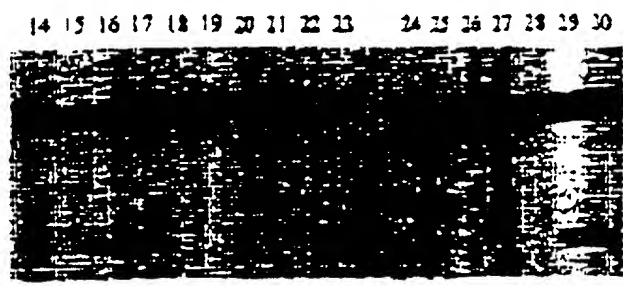
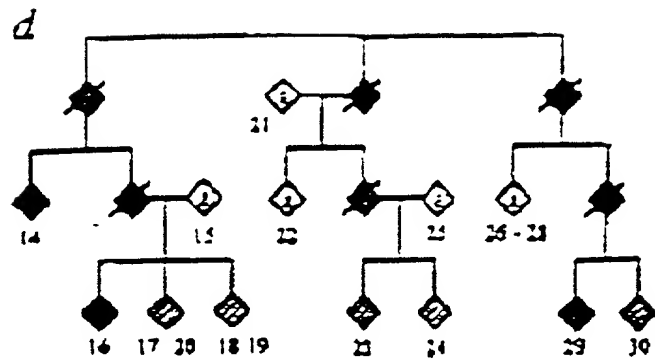


FIG. 3D



← Wild-type

← Mutant

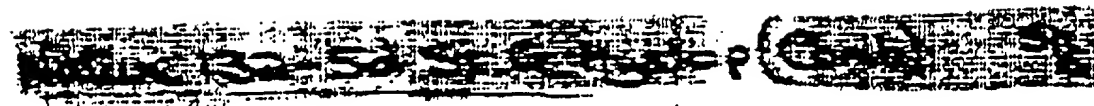


FIG. 3E

Amygdala
 Caudate
 Corpus callosum
 Hippocampus
 Hypothalamus
 Substantia Nigra
 Subthalamic Nucleus
 Thalamus



← 28S

← 18S

FIG. 4

Heart
Brain
Placenta
Lung
Liver
Skeletal Muscle
Kidney
Pancreas

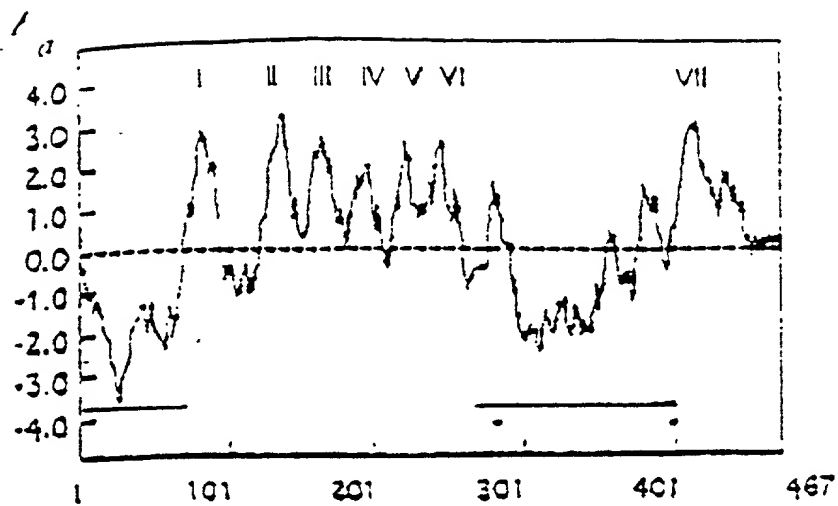


← 28S

← 18S

FIG. 5

A



B

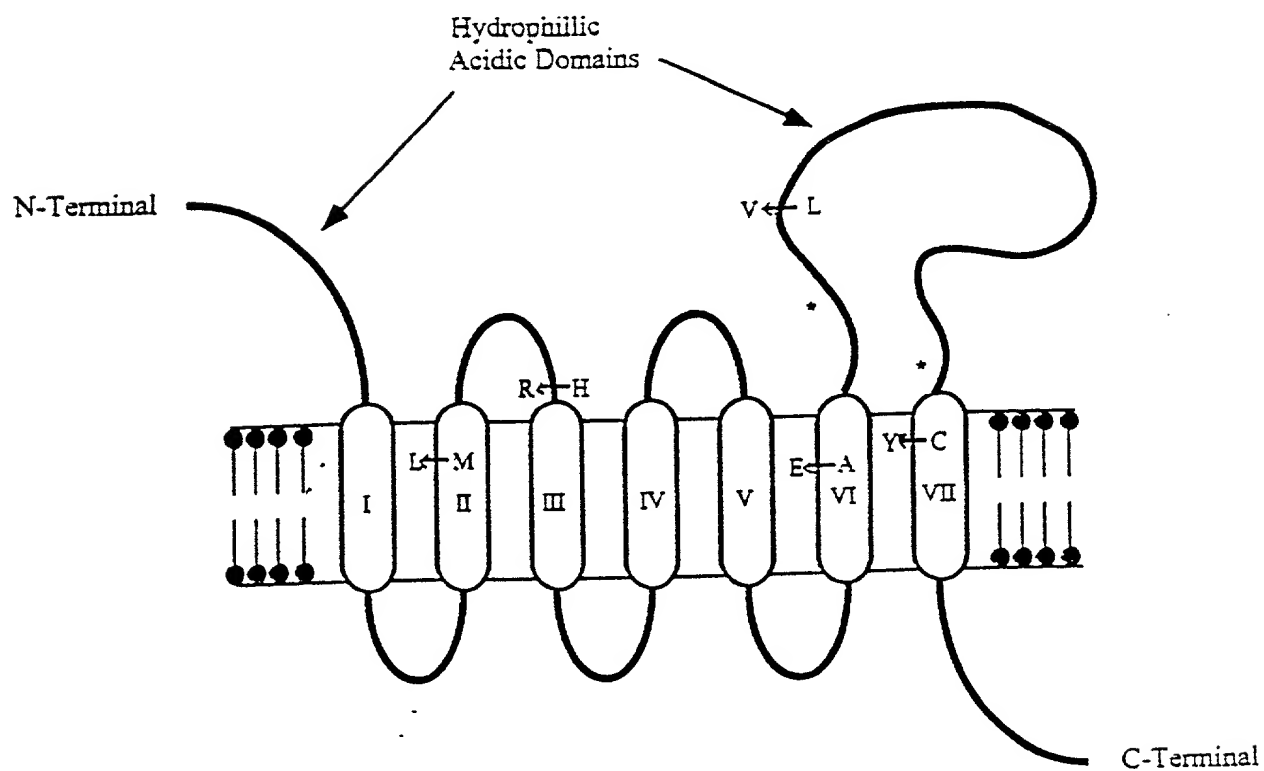


FIG. 6

ES-1

ARMP

ES-1

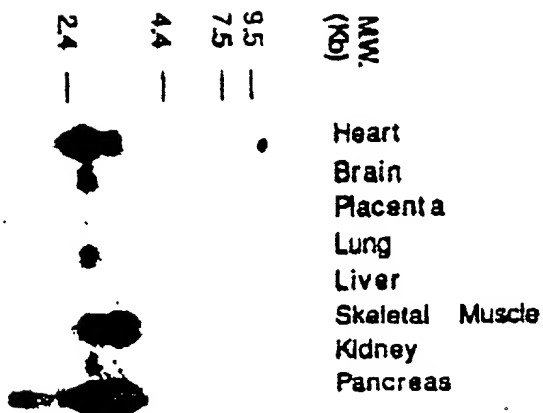
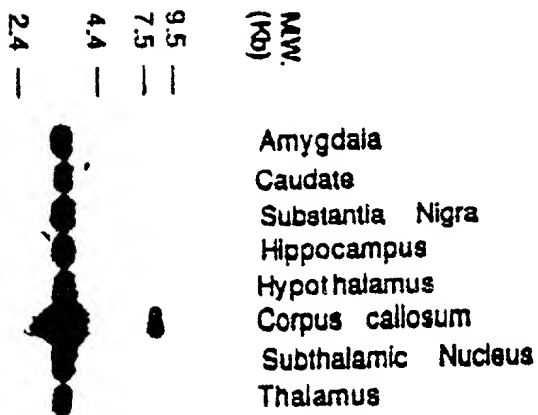
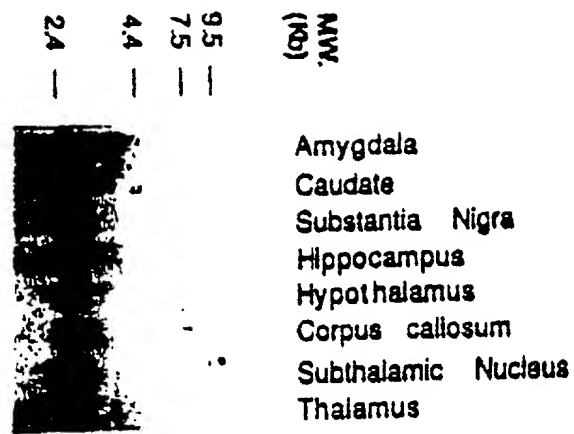


FIG. 7

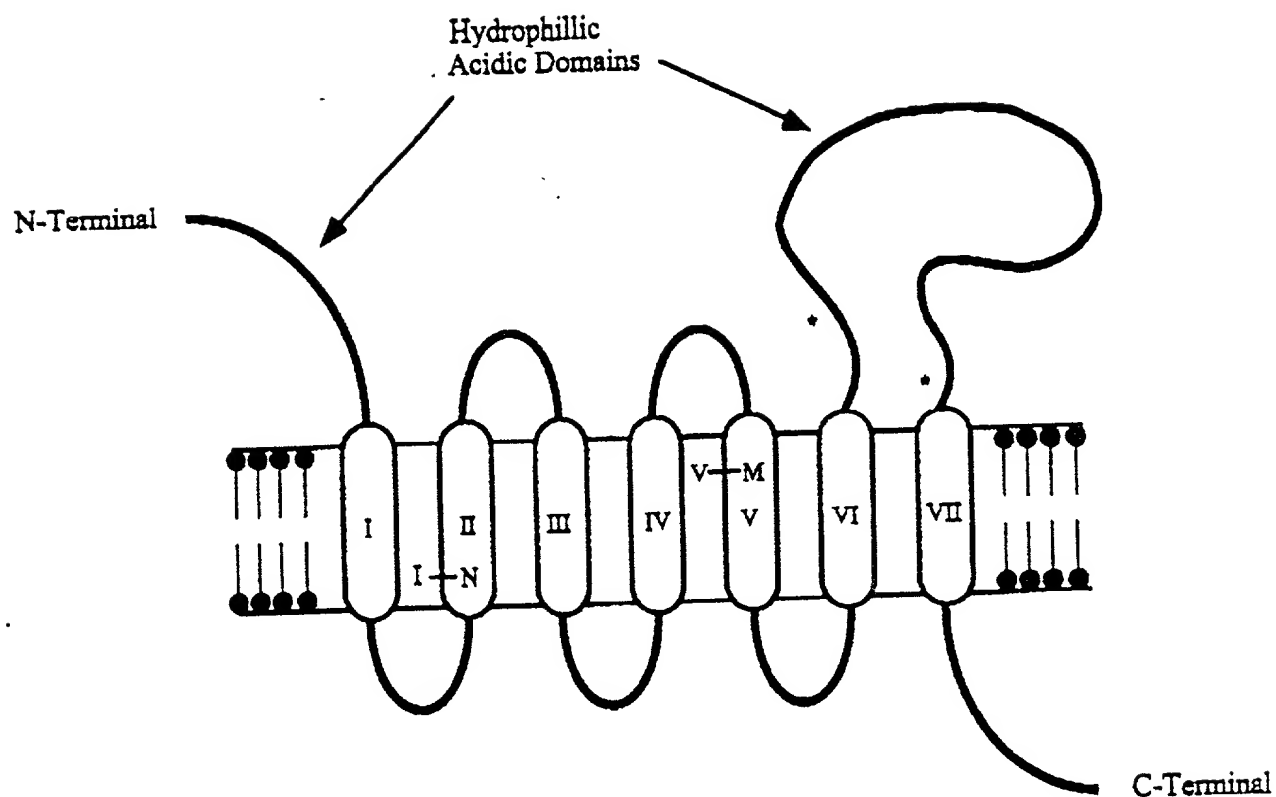


FIG. 8

**COMBINED DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION**

(Original, Design, National Stage of PCT, Supplemental, Divisional, Continuation or CIP)

As a below named inventor, I hereby declare that:

TYPE OF DECLARATION

This declaration is of the following type: (check one applicable item below)

- ☐ original
- ☐ design
- ☐ supplemental
- ☐ national stage of PCT
- ☐ divisional
- ☐ continuation
- ☒ continuation-in-part (CIP)

INVENTORSHIP IDENTIFICATION

My residence, post office address and citizenship are as stated below next to my name, and I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

TITLE OF INVENTION

GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE

SPECIFICATION IDENTIFICATION

the specification of which (check one):

- ☐ is attached hereto.
- ☒ was filed on July 31, 1995 as Application Serial No. 08/509,359 or
- ☐ Express Mail No., as Serial No. not yet known _____ and was amended on (if applicable).
- ☐ was described and claimed in PCT International Application No. _____ filed on _____ and as amended under PCT Article 19 on _____ (if any).

ACKNOWLEDGMENT OF REVIEW OF PAPERS AND DUTY OF CANDOR

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims as amended by any amendment referred to herein.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

PRIORITY CLAIM

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed.

Check one:

- ☒ no such applications have been filed.
- ☐ such applications have been filed as follows:

CLAIM FOR BENEFIT OF EARLIER U.S./PCT APPLICATION(S)

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application.

**PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS
DESIGNATING THE U.S. FOR BENEFIT UNDER 35 USC 120:**

U.S. APPLICATIONS	U.S. FILING DATE	STATUS
08/496,841 (Application Serial No.)	June 28, 1995 (Filing Date)	Pending (Status) (patented, pending, aband.)
08/431,048 (Application Serial No.)	April 28, 1995 (Filing Date)	Pending (Status) (patented, pending, aband.)

POWER OF ATTORNEY

As a named inventor, I hereby appoint the following attorneys and/or agents to prosecute this application and transact all business in the United States Patent and Trademark Office connected therewith:

Steven M. Bauer	Reg. No. 31,481
Paula A. Campbell	Reg. No. 32,503
Joseph A. Capraro, Jr.	Reg. No. 36,471
John J. Cotter	Reg. No. 38,116
Gillian M. Fenton	Reg. No. 36,508
Kerry A. Flynn	Reg. No. 33,693
Duncan A. Greenhalgh	Reg. No. 38,678
Robin D. Kelley	Reg. No. 34,637
Douglas J. Kline	Reg. No. 35,574
John D. Lanza	Reg. No. P-40,060
Robin R. Longo	Reg. No. P-40,071
Edmund R. Pitcher	Reg. No. 27,829
Kurt Rauschenbach	Reg. No. P-40,137
Gretchen A. Rice	Reg. No. 37,429
Christopher W. Stamos	Reg. No. 35,370
Robert J. Tosti	Reg. No. 35,393
Thomas A. Turano	Reg. No. 35,722
Michael J. Twomey	Reg. No. 38,349
Christine C. Vito	Reg. No. 39,061

Direct correspondence to:

Patent Administrator
Testa, Hurwitz & Thibault
High Street Tower
125 High Street
Boston, MA 02110

Direct telephone calls to:

Michael J. Twomey (617) 248-7362

DECLARATION

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

SIGNATURE(S)

Peter H. St. George-Hyslop

Full name of sole or first inventor

Peter H. St. George-Hyslop

Inventor's signature

210 Richview Avenue, Toronto, M5P 3G3

Residence

Same

Post Office Address

Jt. Canadian & U.K.

Citizenship

x Feb 14th 1996

Date

Johanna M. Rommens

Full name of sole or second inventor (if any)

Johanna M. Rommens

Inventor's signature

105 McCaul Street, Toronto M5T 2X4

Residence

Same

Post Office Address

Canadian

Citizenship

x Feb 22, 1996

Date

Paul E. Fraser

Full name of inventor

Paul E. Fraser

Inventor's signature

611 Windermere Avenue, Toronto, Ontario M6S 3L9

Residence

Same

Post Office Address

Canadian

Citizenship

x Feb 21. 96

Date

CERTIFICATE OF FACSIMILE TRANSMISSION:
(703) 308-8724

I hereby certify that this paper is being facsimile
transmitted to the United States Patent and
Trademark Office on the date shown below.

Nicole Roberts
Name of Person signing Certification

Nicole Roberts
Signature

7/23/95
Date

File No.: 1034/1F808-US2

In re Application of: Peter H. ST. GEORGE-HYSLOP, et al.

Serial No: 08/509,359

Examiner: P. Duffy

Filed: July 31, 1995

Group Art Unit: 1645

For: GENETIC SEQUENCES AND PROTEINS RELATED TO ALZHEIMER'S DISEASE

ASSOCIATE POWER OF ATTORNEY

Hon. Commissioner of
Patents and Trademarks
Washington, DC 20231

Sir:

The undersigned attorney of record in the above-identified application hereby appoints:

Michael J. Sweedler (No. 19,937)

S. Peter Ludwig (No. 25,351)

Paul Fields (No. 20,298)

Joseph B. Lerch (No. 26,936)

Melvin C. Garner (No. 26,272)

Ethan Horwitz (No. 27,646)

Beverly B. Goodwin (No. 28,417)

Martin E. Goldstein (No. 20,869)

Adda C. Gogoria (No. 29,714)
Bert J. Lewen (No. 19,407)
Henry Sternberg (No. 22,408)
Peter C. Schechter (No. 31,662)
Robert Schaffer (No. 31,194)
Robert C. Sullivan, Jr. (No. 30,499)
Ira J. Levy (No. 35,587)
Joseph R. Robinson (No. 33,448)
Paul F. Fehlner (No. 35,135)
Phone: (212) 527-7700
Fax: (212) 753-6237

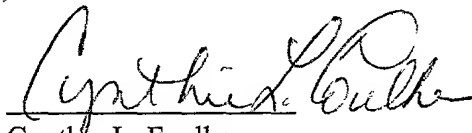
of Darby & Darby, P. C., located at 805 Third Avenue, New York, New York 10022 and

Immac J. Thampoe (No. 36,322)

of Schering-Plough Corporation, 2000 Galloping Hill Road, Kenilworth, New Jersey 07033-0530, as associate attorney with full power to prosecute said application, and to transact all business in the United States Patent and Trademark Office in connection therewith.

Please forward all communications to:

Paul F. Fehlner, Ph.D.
Darby & Darby P.C.
805 Third Avenue
New York, New York 10022


Cynthia L. Foulke
Reg. No. 32,364

Dated: July 22, 1999